A (SEMI-)PARAMETRIC FUNCTIONAL COEFFICIENT LOGARITHMIC AUTOREGRESSIVE CONDITIONAL DURATION MODEL

Marcelo Fernandes
Sao Paulo School of Economics – FGV
School of Economics and Finance, Queen Mary University of London
E-mail: `marcelo.fernandes@fgv.br`

Marcelo C. Medeiros
Department of Economics, Pontifical Catholic University of Rio de Janeiro
E-mail: `mcm@econ.puc-rio.br`

Alvaro Veiga
Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro
E-mail: `alvf@ele.puc-rio.br`

ABSTRACT: In this paper, we propose a class of logarithmic ACD-type models that accommodates overdispersion, intermittent dynamics, multiple regimes, and asymmetries in financial durations. In particular, our functional coefficient logarithmic autoregressive conditional duration (FC-LACD) model relies on a smooth-transition autoregressive specification. The motivation lies on the fact that the latter yields a universal approximation if one lets the number of regimes grows without bound. After establishing sufficient conditions for strict stationarity, we address model identifiability as well as the asymptotic properties of the quasi-maximum likelihood (QML) estimator for the FC-LACD model with a fixed number of regimes. In addition, we also discuss how to consistently estimate a semiparametric variant of the FC-LACD model that takes the number of regimes to infinity. An empirical illustration indicates that our functional coefficient model is flexible enough to model IBM price durations.

KEYWORDS: explosive regimes, neural networks, quasi-maximum likelihood, sieve estimation, smooth transition, stationarity.

ACKNOWLEDGEMENTS: We are grateful to Essie Maasoumi (editor) and an anonymous referee for their constructive comments. We are also indebted to Valentina Corradi, Emmanuel Guerre, Oliver Linton, José António Ferreira Machado, Olivier Scaillet, and Howell Tong for helpful discussions as well as to seminar participants at Queen Mary, SoFiE European Conference (Geneva, Switzerland), Encontro Brasileiro de Econometria (Natal, Brazil), 50 Years of the Econometric Institute (Rotterdam, The Netherlands), Nonlinear Dynamical Methods and Time Series Analysis (Udine, Italy), and Econometrics in Rio (Rio de Janeiro, Brazil) for valuable comments. We thank the CNPq and Pronex/FAPERJ for financial support. The usual disclaimer applies.

1. Introduction

There has recently been a great interest in the implications of price durations in empirical finance. Most analyzes use one of the extensions of the linear autoregressive conditional duration (ACD) model that abound in the literature (Engle and Russell, 1998). Fernandes and Grammig (2006) consider a family of ACD-type models that relies on asymmetric responses to shocks and on a Box-Cox transformation to the conditional duration process. Their family encompasses most ACD-type models in the literature, though there are a few exceptions. Zhang et al. (2001) argue for a nonlinear version based on threshold ACD processes, whereas Meitz and Teräsvirta (2006) propose the smooth transition and the time-varying ACD models. This paper puts forward a novel class of logarithmic ACD-type models based on logistic smooth-transition autoregressive processes. In particular, our functional coefficient logarithmic ACD (FC-LACD) model not only nests the ACD-type processes proposed by Meitz and Teräsvirta (2006) (if formulated in logs), but may also serve as the basis for a semiparametric approach if one lets the number of regimes to grow without bounds.

We first address the theoretical aspects of the FC-LACD process with a fixed number of regimes. In particular, we establish sufficient conditions for strict stationarity and for the existence of higher-order moments. It turns out that the conditions are quite mild in that they do not exclude duration processes with explosive regimes. As in Medeiros and Veiga (2009), we show that explosive regimes may entail very interesting dynamics. In particular, strictly stationary FC-LACD processes with explosive regimes are particularly suitable to model intermittent dynamics: The system spends a large fraction of time in a bounded region, but sporadically develops an instability that grows exponentially for some time and then suddenly collapses.

We then move to establishing sufficient conditions for model identifiability as well as for the existence, consistency, and asymptotic normality of the quasi-maximum likelihood (QML) estimator. We derive consistency and asymptotic normality under first- and second-order moment conditions, respectively. Finally, we develop a sequence of simple Lagrange multiplier (LM) tests to determine the number of limiting regimes. Although we derive the tests using the exponential distribution as reference, we also discuss a robust version so as to cope with non-exponential errors.

We also consider a semiparametric version of the FC-LACD model in which the number of extra regimes M increases with the sample size. The motivation rests on the fact that the logistic smooth transition autoregressive process with $M \to \infty$ acts as a universal neural-network approximation Hornik et al. (1989). The resulting semiparametric model encompasses most first-order ACD-type models in the literature, despite the fact we impose some restrictions on the functional coefficients specification to achieve identification of the nonparametric term as well as to ensure stationarity and geometric ergodicity. To estimate the semiparametric model, we rely on a regularization procedure that penalizes the log-likelihood function as one increases the number of regimes. In particular, we employ the results in Chen and Shen (1998) to provide asymptotic justification for the resulting sieve estimator.

We thus deem that we contribute to the literature in several aspects. First, in contrast to the smooth transition ACD framework of Meitz and Teräsvirta (2006), our FC-LACD specification permits modeling more than two limiting regimes as well as explosive regimes. Second, our framework allows for statistical inference as to what concerns the number of regimes, and hence it is not necessary to impose a priori a certain number of regimes as in Zhang et al. (2001). Third, we also consider the case in which the number of regimes increase with the sample size, so as to obtain a sieve approximation for the conditional duration process. Finally, we demonstrate the practical usefulness of the FC-LACD specification by modeling IBM price durations. The main motivation lies on the fact that early findings clearly reject many of the extant ACD-type specifications in the literature; see Fernandes and Grammig (2006). We show that allowing for multiple regimes facilitates substantially the task of reaching a congruent specification for the IBM price durations.

The remainder of the paper is organized as follows. Section 2 outlines the statistical properties of the FC-LACD process, whereas Section 3 deals with quasi-maximum likelihood estimation for a known fixed number of regimes. Section 4 then proposes a sequential testing procedure to determine the unknown number of regimes. Section 5 considers a semiparametric version of the FC-LACD model by letting the number of regimes increase with the sample size. Section 6 collects the findings of the empirical application. Section 7 summarizes the main results and offers some concluding remarks. We collect all technical details in the Appendix.

2. A FUNCTIONAL COEFFICIENT LOGARITHMIC ACD PROCESS

Let the duration $x_i = t_i - t_{i-1}$ denote the time spell between two events occurring at times t_i and t_{i-1}. We define price duration as the time interval necessary to observe a cumulative change in the mid-price of at least some given value. To account for the serial dependence, we assume that durations follow an accelerated time failure process.

ASSUMPTION 1. *Let $x_i = \psi_i \, \epsilon_i$. The sequence $\{\epsilon_i;\, i \in \mathbb{Z}\}$ of iid random variables has a continuous density function $f > 0$ in $[0, \infty)$, with $\mathbb{E}\left(\epsilon_i \,\middle|\, \mathcal{F}_{i-1}\right) = 1$, where $\mathcal{F}_{i-1}$ is the information set at t_{i-1}. Also, $\psi_i = \mathbb{E}\left(x_i \,\middle|\, \mathcal{F}_{i-1}\right)$ is independent of $\{\epsilon_i, \epsilon_{i+1}, \ldots\}$.*

Assumption 1 is standard; see Drost and Werker (2004). The ACD model of Engle and Russell (1998) assumes a linear specification for the conditional expected duration, viz. $\psi_i = \omega + \alpha \, x_{i-1} + \beta \, \psi_{i-1}$. Bauwens and Giot (2000) propose a logarithmic version of the ACD model with a similar autoregressive structure for the log rather than for the level of the expected duration so as to ensure the positivity of the duration process. In this paper, we propose a more flexible logarithmic ACD-type process.

DEFINITION 1. *x_i follows a functional coefficient logarithmic autoregressive conditional duration (FC-LACD) process with $M+1$ regimes if $x_i = \psi_i \, \epsilon_i$, where ϵ_i and ψ_i satisfy Assumption 1 and*

$$\ln \psi_i = \omega(\ln x_{i-1}) + \alpha(\ln x_{i-1}) \ln x_{i-1} + \beta(\ln x_{i-1}) \ln \psi_{i-1}, \tag{1}$$

$$\begin{aligned}\omega(\ln x_i) &= \omega_0 + \sum_{m=1}^{M} \omega_m\, G_m(\ln x_i), &(2)\\ \alpha(\ln x_i) &= \alpha_0 + \sum_{m=1}^{M} \alpha_m\, G_m(\ln x_i), &(3)\\ \beta(\ln x_i) &= \beta_0 + \sum_{m=1}^{M} \beta_m\, G_m(\ln x_i), &(4)\end{aligned}$$

and

$$G_m(\ln x_i) = G(\ln x_i;\, \gamma_m, c_m) = \frac{1}{1 + \exp\left[-\gamma_m\left(\ln x_i - c_m\right)\right]}. \tag{5}$$

The parameter vector of the FC-LACD model is

$$\boldsymbol{\theta} = (\omega_0, \ldots, \omega_M, \alpha_0, \ldots, \alpha_M, \beta_0, \ldots, \beta_M, c_1, \ldots, c_M, \gamma_1, \ldots, \gamma_M)'\,.$$

The slope parameters γ_m ($m = 1, \ldots, M$) control the smoothness of the regime transitions: e.g., $G_m(\ln x_i)$ converges to a step function as γ_m grows. Equation (5) also implies that log-durations determine the weights at which each regime contributes to the overall dynamics of the process at time t_i. The resulting model thus is quite similar to the self-exciting threshold ACD specification of Zhang et al. (2001). The main differences are that we allow for smooth transitions and that, as in Bauwens and Giot (2000), we model the log rather than the level of the expected duration so as to avoid positivity constraints on the parameter space.

The FC-LACD specification entails several advantages. First, the condition we derive in Subsection 2.1 for strict stationarity does not rule out the presence of explosive regimes. The latter is interesting because it may give way to the moderately high, but very persistent, autocorrelation structure that seems to characterize financial duration data. Second, log transformation aside, our specification nests the smooth transition ACD-type processes in Meitz and Teräsvirta (2006). Third, as in Medeiros and Veiga (2000), one may interpret (2) to (4) as a single-hidden layer neural network with M hidden units. It thus follows that the FC-LACD specification admits a semiparametric variant by letting the number of regimes increase with the sample size. A neural network with a large number of hidden units indeed approximates arbitrarily well any Borel-measurable function; see Hornik et al. (1989).

To establish the statistical properties of the FC-LACD process, we require a standard regularity condition on the error term and on the parameter space.

ASSUMPTION 2. *The error term ϵ_i is such that $\mathbb{E}\,|\ln \epsilon_i| < \infty$ and $\mathbb{E}\,|\epsilon_i|^k < \infty$ for some integer $k \geq 4$.*

ASSUMPTION 3. *The vector $\boldsymbol{\theta}$ is interior to the compact parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^{5M+3}$.*

The asymptotic normality of the QML estimator depends heavily on the fourth-moment requirement in Assumption 2. If the interest lies only on the consistency of the QML estimator, then it suffices to assume that the finiteness of the second moment.

2.1. **Statistical properties: Strict stationarity and geometric ergodicity.** Let $\boldsymbol{u}_i = (\ln\psi_i, \ln\epsilon_i)'$. It is easy to see that $\{\boldsymbol{u}_i;\, i \in \mathbb{Z}\}$ is a Markov chain with homogenous transition probability in view that

$$\boldsymbol{u}_{i+1} = \mathbf{F}\left(\boldsymbol{u}_i;\, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}_{i+1}, \tag{6}$$

where $\mathbf{F}\left(\boldsymbol{u}_i;\, \boldsymbol{\theta}\right) = \left[F\left(\boldsymbol{u}_i;\, \boldsymbol{\theta}\right),\, 0\right]'$ with $F\left(\boldsymbol{u}_i;\, \boldsymbol{\theta}\right) = \omega(\ln x_i) + [\alpha(\ln x_i) + \beta(\ln x_i)]\ln x_i + \alpha(\ln x_i)\ln\epsilon_i$, and $\boldsymbol{\varepsilon}_i = [0,\, \ln\epsilon_i]'$. We are now ready to establish our first theoretical result.

THEOREM 1. *Let x_i follows a FC-LACD process satisfying Assumptions 1 and 2. If $A_0 < 1$, $A_M < 1$, and $A_0\, A_M < 1$, where $A_0 = \alpha_0 + \beta_0$ and $A_M = \sum_{m=0}^{M}\left(\alpha_m + \beta_m\right)$, then strict stationarity and geometric ergodicity hold for the duration process and $\mathbb{E}\left|\ln x_i\right|^k < \infty$, with k as in Assumption 2.*

The sufficient condition in Theorem 1 is intuitive and simple despite not only the highly nonlinear nature of the model but also the extant sufficient conditions in the literature Meitz (2006); Fernandes and Grammig (2006); Meitz and Saikkonen (2008). As in threshold autoregressive models, it suffices to impose constraints only on the two polar regimes. In particular, it allows strictly stationary and ergodic FC-LACD processes to have explosive regimes. This is of particular interest given that such processes are suitable to model intermittent dynamics as in Medeiros and Veiga (2009). An ergodic FC-LACD process with explosive regimes indeed spends a large fraction of time in a bounded region, though it sporadically develops an instability that grows exponentially for some time and then suddenly collapses. As we illustrate in Example 1, even though we only consider first-order specifications, the FC-LACD process admits a highly persistent behavior with moderate values for the autocorrelation function, especially in the presence of explosive regimes.

EXAMPLE 1. *Consider a FC-LACD process with three limiting regimes, exponential errors, and parameters $\omega_0 = 0.005$, $\omega_1 = -0.9$, $\omega_2 = 3$, $\alpha_0 = 0.09$, $\alpha_1 = -0.05$, $\alpha_2 = -0.05$, $\beta_0 = 0.9$, $\beta_1 = 0.6$, $\beta_2 = -0.5$, $\gamma_1 = 1000$, $\gamma_2 = 100$, $c_1 = -2$, and $c_2 = 1$. The condition for strict stationarity holds given that $A_0 = \alpha_0 + \beta_0 = 0.99$, $A_2 = \sum_{m=0}^{2}(\alpha_m + \beta_m) = 0.99$ and $A_0\, A_2 = 0.9801$, despite the explosiveness of the second regime. Figure 1 depicts a simulated path of such duration process and the corresponding autocorrelation function up to the 200th lag.*

3. QUASI-MAXIMUM LIKELIHOOD ESTIMATION

In this section, we address the parametric estimation of the FC-LACD model. To avoid further distributional assumptions, we invoke QML methods.

The derivation of the semiparametric ACD model in Drost and Werker (2004) clarifies that adaptiveness occurs if and only if the error distribution belongs to the two-parameter gamma family with unit mean. It actually turns out that the exponential and gamma scores are proportional, and hence there is no efficiency loss in restricting attention to the exponential distribution. This means that the QML estimator is consistent only if we write the likelihood as if under the assumption of exponential (or standard gamma)

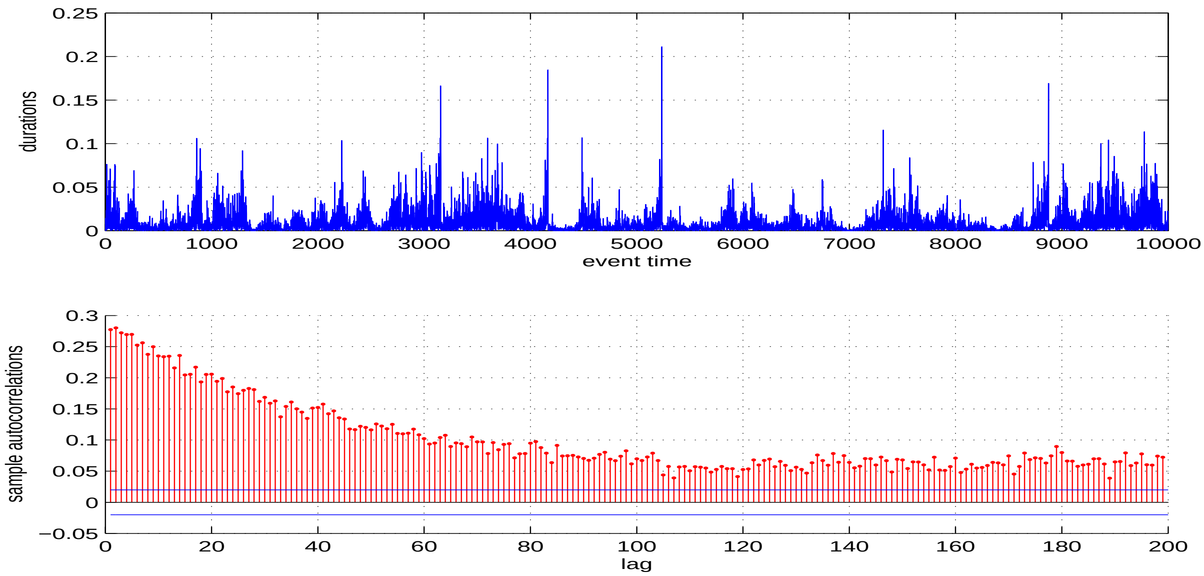

FIGURE 1. Simulated path and autocorrelation function of a FC-LACD process with three limiting regimes, exponential errors, and parameters $\omega_0 = 0.005$, $\omega_1 = -0.9$, $\omega_2 = 3$, $\alpha_0 = 0.09$, $\alpha_1 = -0.05$, $\alpha_2 = -0.05$, $\beta_0 = 0.9$, $\beta_1 = 0.6$, $\beta_2 = -0.5$, $\gamma_1 = 1000$, $\gamma_2 = 100$, $c_1 = -2$, and $c_2 = 1$.

distribution. The quasi-log-likelihood thus reads

$$\mathcal{L}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}), \tag{7}$$

where $\ell_i(\boldsymbol{\theta}) = -\ln\psi_i - \frac{x_i}{\psi_i}$. We treat the unobservable sequence $\{(x_{-i}, \psi_{-i})\,;\, i \in \mathbb{N}\}$ as constant rather than random. The quasi-log-likelihood is thus suitable for practical applications given that it is not conditional on the true initial value (x_0, ψ_0).

To derive the asymptotic properties of the QML estimator, it is convenient to work with the unobserved process $\{(x_{u,i}, \psi_{u,i})\,;\, i \in \mathbb{Z}\}$ as in Lumsdaine (1996), so as to avoid conditioning on initial values. Hence,

$$\begin{aligned} x_{u,i} &= \psi_{u,i}\, \epsilon_{u,i} \\ \ln\psi_{u,i} &= \omega_0 + \alpha_0 \ln x_{u,i-1} + \beta_0 \ln\psi_{u,i-1} + \sum_{m=1}^{M}\Big[\omega_m + \alpha_m \ln x_{u,i-1} + \beta_m \ln\psi_{u,i-1}\Big] G_m(\ln x_i). \end{aligned}$$

The unobserved quasi-log-likelihood conditional on $\mathcal{F}_0 = (x_0, x_{-1}, x_{-2}, \ldots)$ is

$$\mathcal{L}_{u,N}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \ell_{u,i}(\boldsymbol{\theta}), \tag{8}$$

with $\ell_{u,i}(\boldsymbol{\theta}) = -\ln\psi_{u,i} - \frac{x_{u,i}}{\psi_{u,i}}$. Note that (8) establishes the contribution of the last N observations to a log likelihood that we would like to compute if we could observe data from the infinite past. The primary difference between $\mathcal{L}_N(\boldsymbol{\theta})$ and $\mathcal{L}_{u,N}(\boldsymbol{\theta})$ is due to the fact we do not observe $\psi_{u,0}$ and $\{\epsilon_{u,i},\, i \le 0\}$. For the same technical reasons, we also consider the unfeasible QML estimator based on (8). Let

$$\widehat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathsf{argmax}}\, \mathcal{L}_N(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathsf{argmax}}\, \frac{1}{N}\sum_{i=1}^{N} \ell_i(\boldsymbol{\theta}), \tag{9}$$

and

$$\widehat{\boldsymbol{\theta}}_{u,N} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathsf{argmax}}\, \mathcal{L}_{u,N}(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\mathsf{argmax}}\, \frac{1}{N}\sum_{i=1}^{N} \ell_{u,i}(\boldsymbol{\theta}). \tag{10}$$

3.1. **Existence of the QML estimator.** The QML estimator exists only if $\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}\left[\ell_{u,i}(\boldsymbol{\theta})\right]$ exists. The next result immediately follows from Theorem 2.12 in White (1996), which establishes that $\mathcal{L}(\boldsymbol{\theta})$ exists under certain continuity and measurability conditions on the quasi-log-likelihood function.

THEOREM 2. *If the duration x_i follows a strictly stationary and ergodic FC-LACD process with $M+1$ regimes, then, for any parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathcal{L}(\boldsymbol{\theta})$ exists and is finite under Assumptions 1 and 3.*

3.2. **Identifiability of the model.** To carry out statistical inference, we must first show that $\boldsymbol{\theta}_0$ is the unique maximizer of $\mathcal{L}(\boldsymbol{\theta})$. It turns out, however, that we achieve neither global nor local identification of the FC-LACD model without imposing some parametric constraints.

There are three reasons for the model unidentifiability. First, the logistic functions are interchangeable. This means that the empirical loss function of the FC-LACD specification is invariant to regime permutations, and hence there are $(M+1)!$ equal local maxima for the quasi-log-likelihood in (7); see Suarez-Fariñas et al. (2004). Second, the logistic function in (5) is such that $G(\ln x_i;\, \gamma_m, c_m) = 1 - G(\ln x_i;\, -\gamma_m, c_m)$. Third, identifiability also relates to model reducibility. For instance, it is not possible to identify the parameters (γ_m, c_m) if $(\omega_m, \alpha_m, \beta_m)' = \mathbf{0}$, whereas $(\omega_m, \alpha_m, \beta_m, c_m)'$ may take on any value without affecting the value of the quasi-log-likelihood function if $\gamma_m = 0$. We then restrict the parameter space $\boldsymbol{\Theta}$ so as to circumvent such problems.

ASSUMPTION 4. *The parameter space $\boldsymbol{\Theta}$ is such that any vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ satisfies*

C1: $\underline{c} < c_1 < \ldots < c_M < \bar{c}$ *for some finite constants* $\underline{c}$ *and* $\bar{c}$*;*
C2: $\gamma_m > 0$ *for* $m = 1, \ldots, M$ *and*
C3: $(\omega_m, \alpha_m, \beta_m) \neq \mathbf{0}$ *for some* $m \in \{0, \ldots, M\}$.

THEOREM 3. *Assumptions 1 to 4 ensure the global identifiability of the FC-LACD model and that $\mathcal{L}(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0$.*

Despite the fact that Assumption 4 is not verifiable, one may alleviate the risk of irrelevant regimes by carrying out a sequence of LM tests (see Section 4).

3.3. **Asymptotic theory.** To derive the next result, we first establish that the unfeasible QML estimator in (10) converges in probability to $\boldsymbol{\theta}_0$ and then show that the difference between the two QML estimators shrinks to zero as N grows.

THEOREM 4. *Under Assumptions 1 to 4, the QML estimators in (9) and (10) are consistent, i.e.,* $\widehat{\boldsymbol{\theta}}_{u,N} \xrightarrow{p} \boldsymbol{\theta}_0$ *and* $\widehat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}_0$*, for strictly stationary FC-LACD models.*

To complete the asymptotic characterization of the QML estimator, let

$$\mathbf{A}_0 = \mathbb{E}\left[- \left. \frac{\partial^2 \ell_{u,i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}_0} \right] \text{ and } \mathbf{B}_0 = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\left. \frac{\partial \ell_{u,i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}_0} \left. \frac{\partial \ell_{u,i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}_0} \right)$$

and denote their empirical counterparts by

$$\mathbf{A}_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\partial^2 \ln \psi_i}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \left(1 - \frac{x_i}{\psi_i} \right) + \frac{x_i}{\psi_i} \frac{\partial \ln \psi_i}{\partial \boldsymbol{\theta}} \frac{\partial \ln \psi_i}{\partial \boldsymbol{\theta}'} \right] \text{ and}$$

$$\mathbf{B}_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i}{\psi_i} - 1 \right)^2 \frac{\partial \ln \psi_i}{\partial \boldsymbol{\theta}} \frac{\partial \ln \psi_i}{\partial \boldsymbol{\theta}'}.$$

THEOREM 5. *Under the conditions we assume in Theorem 4, it follows that*

$$\sqrt{N}\, (\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}_0^{-1} \mathbf{B}_0 \mathbf{A}_0^{-1} \right) \tag{11}$$

and that $\mathbf{A}_N(\widehat{\boldsymbol{\theta}}_N)$ *and* $\mathbf{B}_N(\widehat{\boldsymbol{\theta}}_N)$ *consistently estimate* $\mathbf{A}_0$ *and* $\mathbf{B}_0$*, respectively.*

4. DETERMINING THE NUMBER OF REGIMES

As the FC-LACD model depends on M, statistical inference must come into play. One solution is to use some model selection criterion. There is a serious drawback in such approach, however. Suppose the data generating mechanism is a FC-LACD process with M regimes. Applying a model selection criterion to decide whether to consider $M + 1$ regimes requires the estimation of an unidentified model with M logistic functions. It thus is impossible to estimate the parameters in a consistent manner, so that numerical problems likely arise in the QML estimation. Besides, the lack of identification also contaminates the likelihood ratio statistic, which does not converge to the usual χ^2 distribution under the null hypothesis of M regimes.

We therefore take a different route. We circumvent the identification problem using sequential LM-type tests. Our sequential testing procedure controls for the significance level of the individual tests using Bonferroni's upper bound for the overall significance level. In what follows, we discuss our framework assuming exponential errors and then show how to robustify the procedure so as to cope with nonexponential errors.

Consider an ergodic FC-LACD process with $M+1$ regimes as in (1)–(5). To test whether it is necessary to include the term corresponding to the $(M+1)$th regime, viz.

$$(\omega_M + \alpha_M \ln x_{i-1} + \beta_M \ln \psi_{i-1})\, G_M(\ln x_{i-1};\, \gamma_M, c_M), \tag{12}$$

we define the null and alternative hypotheses as $\mathbb{H}_M$: $\gamma_M = 0$ and $\mathbb{H}_{M+1}$: $\gamma_M > 0$, respectively. To remedy the lack of identification under the null, we expand the logistic function G_M around $\gamma_M = 0$ as in Meitz and Teräsvirta (2006).

A first-order Taylor expansion of G_M around $\gamma_M = 0$ then yields

$$\begin{aligned} \ln \psi_i = \tilde{\omega}_0 + \tilde{\alpha}_0 \ln x_{i-1} + \tilde{\beta}_0 \ln \psi_{i-1} + \sum_{m=1}^{M-1} \left[\omega_m + \alpha_m \ln x_{i-1} + \beta_m \ln \psi_{i-1}\right] G_m(\ln x_{i-1}) \\ + \delta_1 \ln \psi_{i-1} \ln x_{i-1} + \delta_2 \left(\ln x_{i-1}\right)^2 + O\left(\gamma_M^2\right), \end{aligned} \tag{13}$$

where $\tilde{\omega}_0 = \omega_0 + \frac{1}{2}\,\omega_M - \frac{1}{4}\,\omega_M \gamma_M c_M$, $\tilde{\alpha}_0 = \alpha_0 + \frac{1}{2}\,\alpha_M + \frac{1}{4}\,\gamma_M(\omega_M - \alpha_M c_M)$, $\tilde{\beta}_0 = \beta_0 + \frac{1}{2}\,\beta_M - \frac{1}{4}\,\beta_M \gamma_M c_M$, $\delta_1 = \frac{1}{4}\,\beta_M \gamma_M$, and $\delta_2 = \frac{1}{4}\,\alpha_M \gamma_M$. Under the null of $\mathbb{H}_M$: $\gamma_M = 0$, the specification in (13) reduces to the FC-LACD model with M regimes. Before stating the next result, we first establish some notation. Let $\boldsymbol{\phi} = [\boldsymbol{\theta}', \boldsymbol{\delta}']'$ with $\boldsymbol{\delta} = (\delta_1, \delta_2)'$. The QML estimator of $\boldsymbol{\phi}$ under the null hypothesis of $\mathbb{H}_M$: $\gamma_M = 0$ is $\widehat{\boldsymbol{\phi}}_N = [\widehat{\boldsymbol{\theta}}_N, \mathbf{0}]$. Let $\widehat{\psi}_i \equiv \psi_i\left(\widehat{\boldsymbol{\phi}}_N\right)$ denote the estimate of the expected conditional duration under the null and $\widehat{\boldsymbol{d}}_i = \left.\frac{\partial \ln \psi_i}{\partial \boldsymbol{\phi}}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_N}$ correspond to the derivative of $\ln \psi_i$ with respect to $\boldsymbol{\phi}$ evaluated at the QML estimator $\widehat{\boldsymbol{\phi}}_N$. Although $\widehat{\boldsymbol{d}}_i$ is recursive in that it depends on $\widehat{\boldsymbol{d}}_{i-1}$, it is straightforward to calculate it as a function of the initial value x_0. We are now ready to state the asymptotic distribution of the LM statistic to test $\mathbb{H}_M$: $\gamma_M = 0$ against $\mathbb{H}_{M+1}$: $\gamma_M > 0$.

THEOREM 6. *Let the duration x_i follow a strictly stationary and ergodic FC-LACD process with M regimes. Assumptions 1 to 4 ensure that*

$$LM = N \left[\sum_{i=1}^{N} \left(\frac{x_i}{\widehat{\psi}_i} - 1\right) \widehat{\boldsymbol{d}}_i'\right] \left(\sum_{i=1}^{N} \widehat{\boldsymbol{d}}_i\, \widehat{\boldsymbol{d}}_i'\right)^{-1} \left[\sum_{i=1}^{N} \left(\frac{x_i}{\widehat{\psi}_i} - 1\right) \widehat{\boldsymbol{d}}_i\right] \tag{14}$$

has an asymptotic χ_2^2 distribution under the null of $\mathbb{H}_M$: $\gamma_M = 0$.

To avoid the exponential assumption, one may consider a robust version of the LM test as in Meitz and Teräsvirta (2006), using the tools in Wooldridge (1990). The resulting procedure are as follows.

(1) Estimate the FC-LACD model under the null (with M regimes).
(2) Regress $\frac{\partial}{\partial \boldsymbol{\theta}} \ln \psi_i\big|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_N}$ on $\frac{\partial}{\partial \boldsymbol{\delta}} \ln \psi_i\big|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_N}$ and compute the vector of residuals $\widehat{\boldsymbol{r}}_i$ for $i = 1, \ldots, N$.
(3) Regress 1 on $\left(x_i/\widehat{\psi}_i - 1\right) \widehat{\boldsymbol{r}}_i$ and compute the resulting sum of squared residuals SSR. The robust test statistic $LM_R = N - SSR$ has an asymptotic χ_2^2 distribution under $\mathbb{H}_M$.

We now combine the above statistical ingredients into a coherent modeling strategy. The idea is to test a FC-LACD model with only one regime against an alternative model with two regimes at the significance level λ_1. In the event we reject the null, we keep testing FC-LACD specifications with J regimes against alternative models with $J+1$ regimes at the significance level $\lambda_J = \lambda_1\, C^{1-J}$ for some arbitrary constant $C > 1$. We terminate the testing sequence at the first nonrejection outcome and then estimate the number of extra regimes M of the FC-LACD specification by $\widehat{M} = \bar{J} - 1$, where $\bar{J}$ refers to how many testing runs are necessary to lead to the first nonrejection result. By reducing the significance level at each step, we are able to control the overall significance level and hence to avoid excessively large models. The Bonferroni procedure indeed ensures that such sequence of robust LM tests is consistent and that $\sum_{J=1}^{\bar{J}} \lambda_J$ acts as an upper bound for the overall significance level. As for the selection of the arbitrary constant C, it is good practice to carry out the sequential testing procedure with different values of C.

5. A SEMIPARAMETRIC VARIANT

In this section, we take benefit from the fact that the logistic smooth transition autoregressive specification in (2) to (4) corresponds to a single-hidden layer neural network with M hidden units. This implies that, if M is large enough, it approximates arbitrarily well any Borel-measurable function (Hornik et al., 1989). We therefore consider a semiparametric version of the FC-LACD model in which the number of extra regimes M increases with the sample size. To emphasize the dependence on the sample size, we denote the number of extra regimes by M_N.

DEFINITION 2. *The duration x_i follows a semiparametric FC-LACD process if $x_i = \psi_i\, \epsilon_i$, where ϵ_i and ψ_i satisfy Assumption 1 with $\sigma_\epsilon^2 = \mathbb{E} \ln \epsilon_i^2 \leq \infty$ and*

$$\ln \psi_i = \omega(\ln x_{i-1}) + \beta\, \ln \psi_{i-1} \tag{15}$$

with $|\beta| < 1$ and $\omega(\cdot) < \infty$ belonging to the functional space $\mathcal{H}$ of continuous bounded functions with finite first absolute moments of the Fourier magnitude distributions.

This definition complements well the semiparametric approach of Drost and Werker (2004), whose focus is on the error distribution rather than on the specification of the conditional expected duration. We confine attention to processes that satisfy strict stationarity with finite second moments, geometric ergodicity, and $\beta-$mixingness with exponential decay by assuming that $\omega(\cdot)$ is bounded and that $|\beta| < 1$; see Meitz and Saikkonen (2008). This ensures that we may apply the asymptotic theory in Chen and Shen (1998). In addition, we eliminate the slope functional coefficient in (1) — i.e., $\alpha(z)\, z$ — to achieve identification of the nonparametric component in (15). Finally, we restrict the recursiveness of the conditional expected duration process by assuming that β is constant. Therefore it is possible to rewrite the semiparametric FC-LACD model as a tractable nonlinear model of infinite order, namely,

$$\begin{aligned}\ln x_i &= \ln\psi_i + \ln\epsilon_i = \omega(\ln x_{i-1}) + \beta\,\ln\psi_{i-1} + \ln\epsilon_i \\ &= \sum_{j=0}^{\infty}\beta^j\,\omega(\ln x_{i-1-j}) + \ln\epsilon_i. \qquad (16)\end{aligned}$$

As the largeness of the functional space $\mathcal{H}$ may compromise the estimation, we approximate $\mathcal{H}$ with a sequence $\mathcal{H}_N$ of sieve spaces Grenander (1981); Chen and Shen (1998) that becomes dense in $\mathcal{H}$ as the sample size increases. As the sieve spaces correspond to finite-dimensional parameter spaces, they only require parametric estimation. In particular, we approximate the function $\omega \in \mathcal{H}$ with $\omega_N \in \mathcal{H}_N$, where

$$\omega_N(\cdot) = \omega_0^{(N)} + \sum_{m=1}^{M_N}\omega_m^{(N)}\,G_m^{(N)}(\cdot) \qquad (17)$$

and $G_m^{(N)}(\cdot)$ is the logistic function in (5) with finite parameters $c_m^{(N)}$ and $\gamma_m^{(N)}$. We restrict the parameter space so that $\sum_{m=0}^{M_N}|\omega_m^{(N)}| \le c$, $|c_m^{(N)}| \le c$, and $|\gamma_m^{(N)}| \le c$ for some generic constant $c < \infty$. Makovoz (1996) demonstrates that the approximation error is such that $\|\omega_N - \omega\,\| = O\left(M_N^{-1}\right)$, where $\|\cdot\,\|$ denotes the L^2 norm.

The resulting vector of parameters then is

$$\boldsymbol{\theta}_{(N)} = \left(\omega_0^{(N)},\ldots,\omega_{M_N}^{(N)},\,c_1^{(N)},\ldots,c_{M_N}^{(N)},\,\gamma_1^{(N)},\ldots,\gamma_{M_N}^{(N)},\,\beta\right).$$

Instead of alluding to the sequence $\mathcal{H}_N$ of sieve functional spaces, we sometimes refer to the corresponding sequence $\Theta_{(N)}$ of sieve parameter spaces so as to emphasize the parametric nature of the estimation problem. In accordance with the sieve literature, we then approximate the first term of the right-hand side of (16) by

$$\ln\psi_i^{(N)} = \sum_{j=0}^{J_N}\beta^j\,\omega_N(\ln x_{i-1-j}). \qquad (18)$$

This means that we actually employ two sieve approximations. The first truncates the infinite summation in (16) by means of J_N, whereas the second relates to the finite number of extra regimes M_N in the neural network. The next result documents the conditions under which our semiparametric approach is consistent. The proof is straightforward, relying on the fact that (18) converges to the first term of the right-hand side of (16) as both J_N and M_N go to infinity with the sample size.

THEOREM 7. *If the duration process x_i satisfies Definition 2, the sieve approximation error is negligible as long as $J_N \to \infty$ and $M_N^3 \ln M_N = O(N)$.*

To avoid overfitting, we take a regularization approach by penalizing the empirical loss function. Let

$$\mathcal{L}_N(\boldsymbol{\theta}_{(N)}) = \frac{1}{N}\sum_{i=1}^{N}\ell_i^{(N)}\left(\boldsymbol{\theta}_{(N)}\right), \qquad (19)$$

where

$$\ell_i^{(N)}\left(\boldsymbol{\theta}_{(N)}\right) = -\ln \psi_i^{(N)} - \frac{x_i}{\psi_i^{(N)}} + \lambda_N \left\|\boldsymbol{\theta}_{(N)}\right\|,$$

λ_N is a regularization factor that shrinks to zero as the sample size increases. The sieve estimator then is

$$\widehat{\boldsymbol{\theta}}_{(N)} = \underset{\boldsymbol{\theta}_{(N)} \in \boldsymbol{\Theta}_{(N)}}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \ell_i^{(N)}\left(\boldsymbol{\theta}_{(N)}\right). \tag{20}$$

Given that $\mathcal{L}\left(\boldsymbol{\theta}_{(N)}\right) = \mathbb{E}\left[\ell_i^{(N)}\left(\boldsymbol{\theta}_{(N)}\right)\right]$ is uniquely identified, the sieve estimator in (20) is well defined and hence the results in Chen and Shen (1998) hold.

6. Revisiting IBM price durations

In this section, we estimate both the parametric and semiparametric versions of the FC-LACD model for the price durations of the IBM stock traded on the New York Stock Exchange (NYSE) from September to November 1996. In contrast to the empirical analysis of IBM durations in Zhang et al. (2001), we do not fix the number of regimes in that we let the data determine the proper number of regimes. In particular, we employ a sequence of LM-type tests to choose the number of regimes in the parametric specification, whereas we use a regularization approach that penalizes the likelihood as the number of sieves increases in the semiparametric context. We define price duration as the time interval necessary to observe a cumulative change in the mid-price of at least \$0.125. The main interest in models for price durations is due to the fact that they permit retrieving intraday estimates of the instantaneous volatility of the price process Engle and Russell (1998).

Apart from the opening auction, NYSE trading is continuous from 9:30 to 16:00. We remove all durations between events recorded outside the regular opening hours of the NYSE as well as overnight spells. It is well known that financial durations feature a strong time-of-the-day effect. We therefore consider diurnally adjusted durations $x_i = D_i/\varrho(t_i)$, where D_i is the plain price duration in seconds and $\varrho(\cdot)$ denotes the diurnal factor determined by first averaging the durations over thirty minutes intervals for each day of the week and then fitting a cubic spline with nodes at each half hour. The resulting (diurnally adjusted) durations serve as input for the remainder of the analysis.

A comparison between price and trade durations mirrors the fact that the IBM stock is very liquid. In particular, more than 75% of the trade durations do not exceed 30 seconds and it takes several transactions to alter the mid-quote price by at least \$0.125. The sample size indeed reduces from 60,454 to 6,728 observations once we filter the data to compute price durations. Table 1 describes the main statistical properties of the IBM price durations. We compute descriptive statistics for both plain and diurnally adjusted data for two subsamples. We employ the first subsample with 4,484 observations for estimation purposes, reserving the second subsample with 2,244 observations for out-of-sample analysis.

The distributions of the price durations in the first and second subsamples are substantially different, regardless of the time-of-the-day adjustment. For instance, if one restricts attention to the diurnally

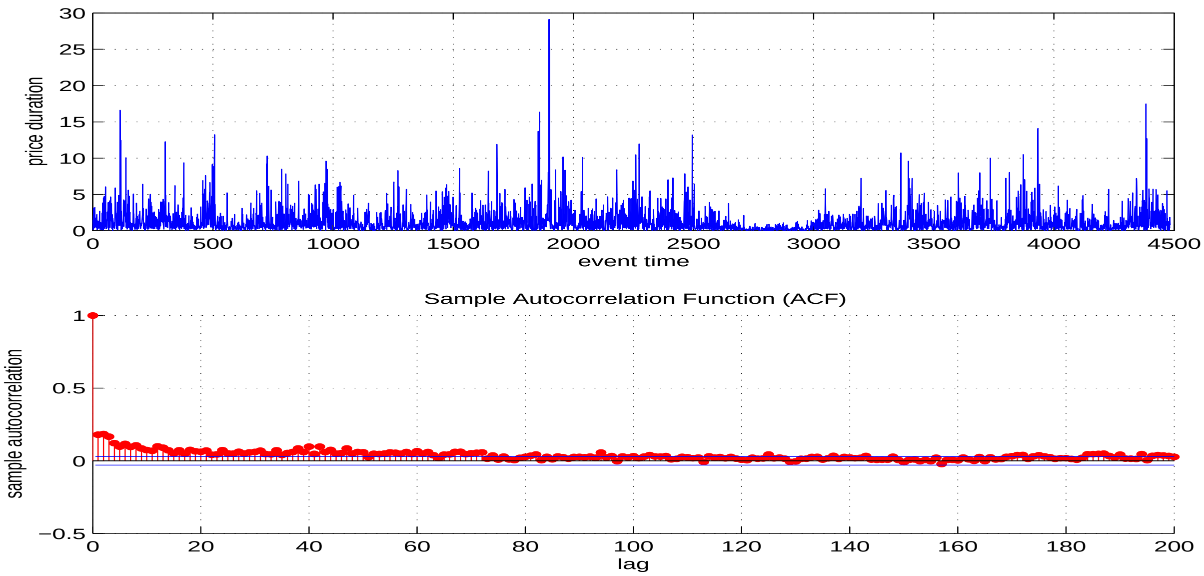

FIGURE 2. The first plot displays the time series of IBM price durations from September to November 1996, whereas the second plot exhibits its sample autocorrelation function up to 200 lags. The data correspond to diurnally adjusted durations.

adjusted series, the first-subsample mean, standard deviation, first quartile, and median are about twofold their counterparts in the second subsample. In addition, the third quartile declines by more than one half from the first to the second subsample, whereas the maximum value in the first subsample is threefold the maximum in the second subsample. The minimum value and overdispersion are the only statistics that remain approximately constant across subsamples.

The evidence in favor of overdispersion is also robust to the time-of-the-day effect. The latter feature ensures that it is not an artifact due to data seasonality. Figure 2 displays the diurnally adjusted series of IBM price durations as well as its sample autocorrelation function up to 200 lags. It reveals that IBM price durations are very persistent in that there are significant positive values in the sample autocorrelation function at very high orders. Altogether, the combination of overdispersion and persistent autocorrelation in IBM price durations warrants the estimation of FC-LACD models with multiple regimes.

We first estimate by quasi-maximum likelihood the FC-LACD model of first order using the exponential distribution as reference. Table 2 reports the estimation and testing results for models with one and two regimes given that our modeling cycle strategy indicates that IBM price durations feature only two limiting regimes. The LM test for additional regimes indeed does not reject the null of only two limiting regimes at the usual levels of significance. Although the transition between the two regimes is very abrupt given the large value of $\hat{\gamma}_1$, Figure 3 shows that there are enough observations (i.e., about 200 data points) within the transition phase to estimate the parameters of the logistic function with reasonable precision.

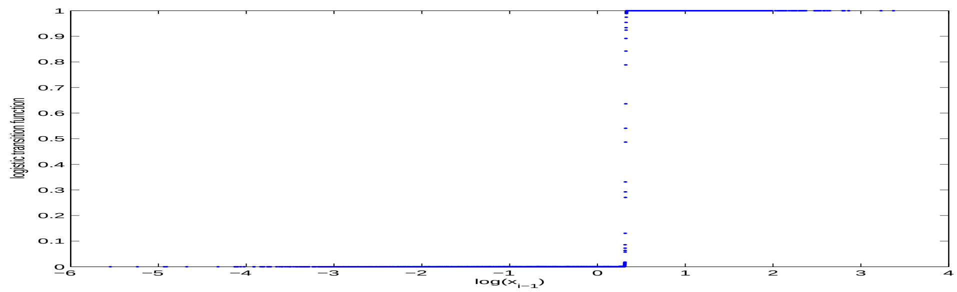

FIGURE 3. The graph plots the logistic function against the sample values of the transition variable for $c_1 = 0.3210$ and $\gamma = 496.99$. The transition variable is the past value of the logarithm of the diurnally adjusted IBM price duration. The sample consists of the first 4,484 observations from the period ranging from September to November 1996.

The first regime is extremely persistent, with $\hat{A}_0 = \hat{\alpha}_0 + \hat{\beta}_0 = 0.9909$, whereas persistence subsides in the second regime given that $\hat{A}_1 = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\beta}_0 + \hat{\beta}_1 = 0.8696$. The less persistent second regime mostly affects larger durations in view that $\exp(\hat{c}_1) = 1.3784$ lies slightly above the sample mean of the IBM price durations, at the 78% percentile. This is somewhat in line with the evidence put forth by Zhang et al. (2001), though they assume nonsmooth transitions between three fixed (rather than estimated) regimes in their threshold ACD model.

The results for the FC-LACD model with one regime, which corresponds to the logarithmic ACD model of Bauwens and Giot (2000), show that ignoring the second regime affects substantially the analysis. The persistence of the one-regime model is a convex combination of the very distinct degrees of persistence that characterize the first and second regimes of the FC-LACD model. In particular, it is closer to the persistence in the first regime, which seems to prevail for 3,117 out of the 4,484 observations of the in-sample period. Allowing for the second regime not only entails a better picture of the persistent nature of IBM price durations, but also substantially improves both the in-sample and out-of-sample fits as measured by the quasi-likelihood function values.

Figure 4 displays the actual and fitted values of the IBM price duration for the full sample. Although the fitted values are not as volatile as the realizations of the IBM price durations, it is evident that it tracks well the movements in the latter series. The in-sample and out-of-sample correlations between the actual and fitted log-values are quite reasonable, namely, 0.3832 and 0.3069, respectively. They add up to an overall correlation between actual and fitted log-values of 0.4391 in the full sample. Furthermore, the in-sample and out-of-sample residuals of the FC-LACD model with two regimes also have well-behaved distributions in that their mean and standard deviation are close to unity (as expected given the exponential benchmark). The in-sample residuals have a mean of 1.0001 with a standard deviation of 1.1536, whereas the mean and standard deviation of the out-of-sample residuals are 0.8949 and 1.0152, respectively. The overdispersion coefficients of the in-sample and out-of-sample residuals are respectively 1.1536 and 1.1344, and hence well below the overdispersion that we report in Table 1.

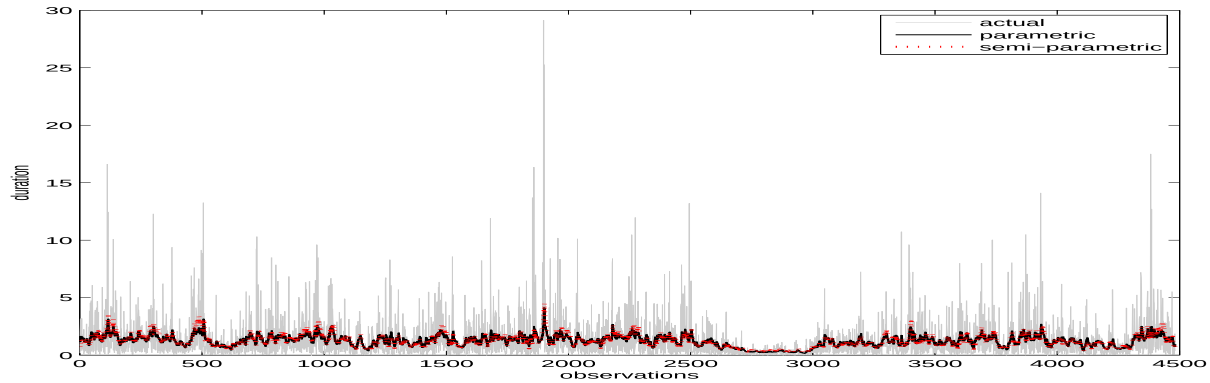

FIGURE 4. The plot displays the actual and fitted values of the IBM price durations from September to November 1996. Actual values are in gray, corresponding to diurnally adjusted IBM price durations. Fitted values are in black for the FC-LACD model with two regimes and in red for the semiparametric variant.

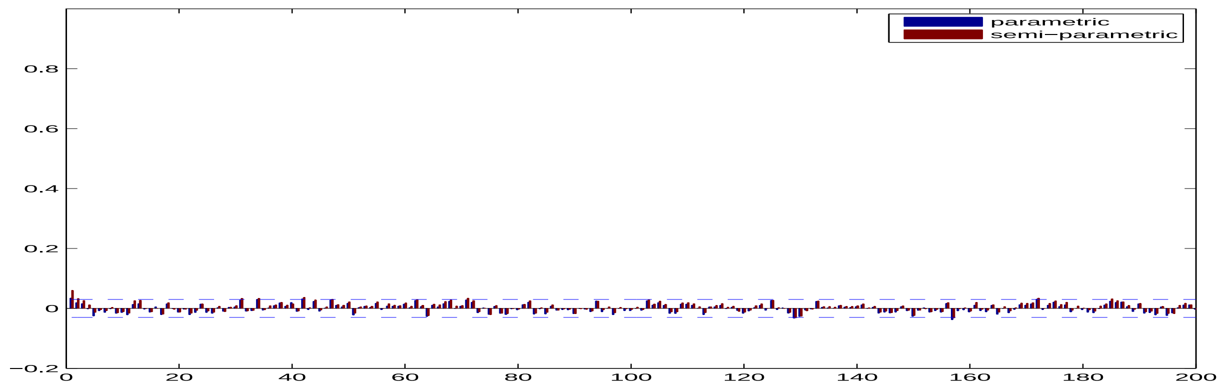

FIGURE 5. The plot displays the autocorrelation function values up to 200 lags of the in-sample parametric and semiparametric residuals (blue and red bars, respectively).

To check for misspecification, we inspect whether the in-sample and out-of-sample residuals display any serial correlation by looking at the sample autocorrelation function up to 200 lags. Table 2 documents that the FC-LACD model with two regimes does a much better job in accounting for the serial dependence in the IBM price durations than the logarithmic ACD model of Bauwens and Giot (2000). Figure 5 shows that the residual autocorrelation reduces by a palpable amount as one allows for the second regime. The decline is particularly strong for the in-sample residuals in that their maximum autocorrelation (in magnitude) for the one-regime model is about twofold the one of the FC-LACD model with two regimes.

We next turn our attention to the semiparametric FC-LACD. Theorem 7 dictates that $M_N^3 \ln M_N = O(N)$, which yields $15 < M_N < 16$ for $N = 4,484$ if we set the constant to one in the order of magnitude. It turns out that varying the maximum number of regimes from 10 to 20 does not change the results, not even minimally, because the regularization approach eliminates most regimes. As for the latter, we play around with the shrinkage parameter λ_N in (19) to assess the robustness of the estimation results. We report the estimates only for $\lambda_N = \sqrt{\frac{\ln T \ln p}{T}} = 0.16$, where p is the number of parameters in the model, even if we also contemplate several alternative degrees of regularization (e.g., $\lambda_N = 1/\ln T$ or $\lambda_N = \ln T/T$). More specifically, the estimates are very similar for any λ_N between 0.01 and 1 essentially because the regularization yields a specification that, in the overall, resembles a lot the parametric model with two regimes.

This is interesting because, in this case, the parametric FC-LACD model 'nests' the semiparametric specification by letting not only ω, but also α and β to change with the past duration. It is thus not surprising that the fit worsens as we move from the parametric to the semiparametric model. The in-sample and out-of-sample correlations between the actual and fitted log-values remain nonetheless quite reasonable, at 0.3763 and 0.3092, respectively. In addition, the mean of the in-sample residuals is 1.0007, with a standard deviation of 1.1755, whereas the mean and standard deviation of the out-of-sample residuals are 0.9069 and 1.0823, respectively. Finally, the semiparametric residuals exhibit a little bit more of autocorrelation than the parametric residuals (see Figure 5). As before, this is not surprising given that the semiparametric specification ends up restricting α to zero and β to 0.8930 across all regimes. Note that the semiparametric estimate of β is virtually the same of the beta estimate we obtain for the LACD model in Table 2. All in all, the semiparametric results seems, interestingly, to vouch for the parametric FC-LACD specification with two regimes.

7. CONCLUSION

This paper proposes a functional coefficient LACD model that accommodates overdispersion, intermittent dynamics, multiple regimes, as well as sign and size asymmetries in financial durations. In particular, we rely on a very flexible smooth-transition autoregressive specification. The motivation lies on the fact that it gives way to a semiparametric version of the model as the number of regimes goes to infinity. We formally address how to consistently estimate the parametric FC-LACD model with fixed number of regimes by quasi-maximum likelihood as well as the semiparametric counterpart.

An empirical illustration indicates that our nodel is flexible enough to model IBM price durations in a congruent manner. This is in stark contrast with the alternative model with a single regime, whose residuals display much larger autocorrelations.

APPENDIX A. PROOFS

Proof of Theorem 1. We start by casting the FC-LACD process with multiple regimes into a smooth transition autoregressive moving average (STARMA) representation. Let $\bar{\omega}_{i-1} \equiv \omega(\ln x_{i-1})$, $\bar{\alpha}_{i-1} \equiv \alpha(\ln x_{i-1})$, and $\bar{\beta}_{i-1} \equiv \beta(\ln x_{i-1})$. It follows from (1) that the duration process has the following STARMA(1,1) representation:

$$\ln x_i = \bar{\omega}_{i-1} + \left[\bar{\alpha}_{i-1} + \bar{\beta}_{i-1}\right] \ln x_{i-1} + \ln \varepsilon_i - \bar{\beta}_{i-1} \ln \varepsilon_{i-1}. \tag{A.1}$$

Following similar steps to Zhang et al. (2001), it is straightforward to show that the Markov chain in (6) is a ϕ-irreducible T-chain. This means that we may apply the drift criterion of Tweedie (1975) to derive sufficient conditions for strict stationarity and geometric ergodicity. In addition, Theorem 4.1 in Ling (1999) implies that both strict stationarity and ergodicity of the functional coefficient ARMA model depends exclusively on its autoregressive part, and hence we confine attention to the analogous STAR(1) process with $M+1$ regimes

$$y_i = \bar{\nu}_{i-1} + \bar{\zeta}_{i-1}\, y_{i-1} + \varsigma_i, \tag{A.2}$$

where $\bar{\nu}_{i-1} \equiv \nu_0 + \sum_{m=1}^{M} \nu_m\, G_m(y_i)$, $\bar{\zeta}_{i-1} \equiv \zeta_0 + \sum_{m=1}^{M} \zeta_m\, G_m(y_{i-1})$, and the error term ς_i is iid with $\mathbb{E}\,|\varsigma_i| < \infty$. The sufficient conditions for strict stationarity that we derive are exactly the same for TAR(1) processes; see, e.g., Chen and Tsay (1991), though our derivation differs in view that (A.2) involves smooth transitions. For any $e_C > 0$, there exists a positive constant $C > \max\{|\,\underline{c}\,|, |\,\bar{c}\,|\}$ such that $\left|\bar{\zeta}_{i-1} - \zeta_0\right| \leq e_C$ for any $y_{i-1} < -C$ and $\left|\bar{\zeta}_{i-1} - \sum_{m=0}^{M} \zeta_m\right| \leq e_C$ for any $y_{i-1} > C$. It then follows that

$$y_i = \bar{\nu}_{i-1} + \mathbf{1}_{\{y_{i-1}<-C\}}\, \bar{\zeta}_{i-1}\, y_{i-1} + \mathbf{1}_{\{|y_{i-1}|\leq C\}}\, \bar{\zeta}_{i-1}\, y_{i-1} + \mathbf{1}_{\{y_{i-1}>C\}}\, \bar{\zeta}_{i-1}\, y_{i-1} + \varsigma_i,$$

where $\mathbf{1}_A$ is the indicator function that takes value one if A is true and zero, otherwise. Taking absolute values of both sides gives way to

$$\begin{aligned} |y_i| &\leq L_C + \mathbf{1}_{\{y_{i-1}<-C\}} \left|\bar{\zeta}_{i-1}\right| |y_{i-1}| + \mathbf{1}_{\{y_{i-1}>C\}} \left|\bar{\zeta}_{i-1}\right| |y_{i-1}| + |\varsigma_i| \\ &\leq L_C + \zeta^+_{i-1}\, |y_{i-1}| + |\varsigma_i| \leq |y_0| \textstyle\prod_{j=0}^{i-1} \zeta^+_j + \displaystyle\sum_{k=1}^{i-1} \left(|\varsigma_k| + L_C\right) \textstyle\prod_{j=k}^{i-1} \zeta^+_j + |\varsigma_i| + L_C, \end{aligned}$$

where $\zeta^+_{i-1} \equiv \mathbf{1}_{\{y_{i-1}<-C\}} \left(|\zeta_0| + e_C\right) + \mathbf{1}_{\{y_{i-1}>C\}} \left(\left|\sum_{m=0}^{M} \zeta_m\right| + e_C\right)$ and L_C is a positive constant that exceeds $|\bar{\nu}_{i-1}| + \mathbf{1}_{\{y_{i-1}<-C\}} \left|\bar{\zeta}_{i-1}\right| C$. We then take conditional expectation yielding

$$\begin{aligned} \mathbb{E}\left(|y_i| \,\middle|\, y_0\right) &\leq |y_0|\, \mathbb{E}\left(\textstyle\prod_{j=0}^{i-1} \zeta^+_j \,\middle|\, y_0\right) + \displaystyle\sum_{k=1}^{i-1} \mathbb{E}\left[\left(|\varsigma_k| + L_C\right) \textstyle\prod_{j=k}^{i-1} \zeta^+_j \,\middle|\, y_0\right] + \mathbb{E}\,|\varsigma_i| + L_C \\ &= |y_0|\, \mathbb{E}\left(\textstyle\prod_{j=0}^{i-1} \zeta^+_j \,\middle|\, y_0\right) + L^*_C \left[1 + \displaystyle\sum_{k=1}^{i-1} \mathbb{E}\left(\textstyle\prod_{j=k}^{i-1} \zeta^+_j \,\middle|\, y_0\right)\right], \end{aligned}$$

where $L^*_C \equiv \mathbb{E}\,|\varsigma_1| + L_C$. We now have four cases to evaluate according to the signs of ζ_0 and $\zeta_* \equiv \sum_{m=0}^{M} \zeta_m$. In the first case, we consider $\zeta_0 > 0$ and $\zeta_* > 0$. It then holds that

$$\mathbb{E}\left(|y_1| \,\middle|\, y_0\right) \leq |y_0|\, \mathbb{E}\left(\zeta^+_0 \,\middle|\, y_0\right) + L^*_C \leq |y_0| \left[\mathbf{1}_{\{y_0<-C\}} \left(|\zeta_0| + e_C\right) + \mathbf{1}_{\{y_0>C\}} \left(|\zeta_*| + e_C\right)\right] + L^*_C,$$

and hence $\mathbb{E}\left(|y_1| \,\middle|\, y_0 < -C\right) \leq |y_0| \left(|\zeta_0| + e_C\right) + L_C$.

If $0 < \zeta_0 < 1$, it is always possible to choose $e_C < 1 - |\zeta_0|$, so that the drift criterion in Tweedie (1975) holds. Analogously, $\mathbb{E}\left(|y_1| \,\middle|\, y_0 > C\right) \leq |y_0| \left(|\zeta_*| + e_C\right) + L_C$, and so the same result follows if $0 < \zeta_* < 1$. In the second case, we assume that $\zeta_0 < 0$ and $\zeta_* < 0$. It then follows that $\mathbb{E}\left(|y_2| \,\middle|\, y_0\right) \leq |y_0|\, \mathbb{E}\left(\zeta^+_1\, \zeta^+_0 \,\middle|\, y_0\right) + L^*_C \left[1 + \mathbb{E}\left(\zeta^+_1 \,\middle|\, y_0\right)\right]$, where

$$\mathbb{E}\left(\zeta^+_1\, \zeta^+_0 \,\middle|\, y_0\right) = \Pr\left(y_1 < -C \,\middle|\, y_0 < -C\right) \left(|\zeta_0| + e_C\right)^2 + \Pr\left(y_1 > C \,\middle|\, y_0 < -C\right) \left(|\zeta_0| + e_C\right) \left(|\zeta_*| + e_C\right). \tag{A.3}$$

However, for any $a_C > 0$, there exists some constant C that bounds from above the first term of the right-hand side of (A.3). This means that $\mathbb{E}\left(\zeta_1^+ \zeta_0^+ \,\middle|\, y_0\right) \le (|\zeta_0| + e_C)(|\zeta_*| + e_C) + a_C = \zeta_0\zeta_* + (|\zeta_0| + |\zeta_*|)\, e_C + e_C^2 + a_C$ satisfies the criterion in Tjøstheim (1990) (i.e., it does not exceed one) if $\zeta_0\zeta_* < 1$ given that both e_C and a_C are arbitrarily small. As before, the same reasoning applies to the case in which $y_0 > C$, yielding exactly the same condition. Finally, the third and fourth cases are symmetrical and hence we consider only the case of $\zeta_0 < 0$ and $0 < \zeta_* < 1$. Letting $h \equiv \inf_{i\in\mathbb{Z}_+} \left|\zeta_0\, \zeta_*^{i-1}\right| < 1$, observe that

$$\mathbb{E}\left(|y_h| \,\middle|\, y_0\right) \quad \le \quad |y_0|\, \mathbb{E}\left(\prod_{j=0}^{h-1} \zeta_j^+ \,\middle|\, y_0\right) + L_C^* \left[1 + \sum_{k=1}^{h-1} \mathbb{E}\left(\prod_{j=k}^{h-1} \zeta_j^+ \,\middle|\, y_0\right)\right].$$

The argument $\prod_{j=0}^{h-1} \zeta_j^+$ will differ from zero only for the paths $\left\{\zeta_0^+, \ldots, \zeta_{h-1}^+\right\}$ whose values are all greater than C in magnitude. To avoid a burdensome notation, we denote these paths by $\mathcal{P}_j$, with $j = 1, \ldots, 2^h$. Therefore,

$$\mathbb{E}\left(\prod_{j=0}^{h-1} \zeta_j^+ \,\middle|\, y_0\right) = \sum_{j=1}^{2^h} (|\zeta_0| + e_C)^{a_j}\, (|\zeta_*| + e_C)^{b_j}\, \Pr\left(\mathcal{P}_j \,\middle|\, y_0\right),$$

where $a_j \equiv \sum_{k=1}^h \mathbf{1}_{\{y_{h-k} < -C\}}$ and $b_j \equiv \sum_{k=1}^h \mathbf{1}_{\{y_{h-k} > C\}}$. As before, it is straightforward to show that, for $|y_0| > C$, the probability of $\{y_\ell < -C\}$ is arbitrarily small for any $\ell = 1, \ldots, h-1$ and the criterion in Tjøstheim (1990) depends exclusively on the values of $(|\zeta_0| + e_C)$ and $(|\zeta_*| + e_C)$. It indeed turns out that $\mathbb{E}\left(\prod_{j=0}^{h-1} \zeta_j^+ \,\middle|\, y_0\right) < 1$ for any $0 < \zeta_* < 1$ such that $\left|\zeta_0\, \zeta_*^{h-1}\right| < 1$. *Q.E.D.*

Proof of Theorem 2. The model given by (1)–(5) is continuous in the parameter vector $\boldsymbol{\theta}$ given that, for any value of $\ln x_i$, the logistic function in (5) depends in a continuous manner on γ_m and c_m. Similarly, the model is also continuous in $\ln x_i$, and hence measurable for any fixed value of the parameter vector $\boldsymbol{\theta}$. The stationarity condition of Theorem 1 then ensures that $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ln|\psi_{u,i}|\right]$ is finite, and thus $\mathbb{E}\,|\ell_{u,t}(\boldsymbol{\theta})| < \infty$ for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. *Q.E.D.*

Proof of Theorem 3. Let $\mathbf{z}_i \equiv [1, \ln x_{i-1}, \ln\psi_{i-1}]'$, $\boldsymbol{\phi}_j \equiv [\omega_j, \alpha_j, \beta_j]'$ for $j = 0, \ldots, M$ and $\boldsymbol{\rho}_m \equiv (\gamma_m, c_m)'$ for $m = 1, \ldots, M$. The parameter vector is $\boldsymbol{\theta} = \left[\boldsymbol{\phi}_0', \ldots, \boldsymbol{\phi}_M', \boldsymbol{\rho}_1', \ldots, \boldsymbol{\rho}_M'\right]'$. Consider now another parameter vector $\widetilde{\boldsymbol{\theta}} \equiv \left[\widetilde{\boldsymbol{\phi}}_0', \ldots, \widetilde{\boldsymbol{\phi}}_M', \widetilde{\boldsymbol{\rho}}_1', \ldots, \widetilde{\boldsymbol{\rho}}_M'\right]'$ such that

$$\boldsymbol{\phi}_0'\mathbf{z}_i + \sum_{m=1}^{M} \boldsymbol{\phi}_m'\mathbf{z}_i\, G(\ln x_{i-1}; \boldsymbol{\rho}_m) = \widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_i + \sum_{m=1}^{M} \widetilde{\boldsymbol{\phi}}_m'\mathbf{z}_i\, G(\ln x_{i-1}; \widetilde{\boldsymbol{\rho}}_i). \tag{A.4}$$

To show global identifiability of the FC-LACD model, we must demonstrate that Assumption 4 ensures that (A.4) holds if and only if $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$. It follows from (A.4) that

$$\boldsymbol{\phi}_0'\mathbf{z}_i - \widetilde{\boldsymbol{\phi}}_0'\mathbf{z}_i - \sum_{j=1}^{2M} \bar{\boldsymbol{\phi}}_j'\mathbf{z}_i\, G\left(\ln x_{i-1}; \bar{\boldsymbol{\rho}}_j\right) = 0, \tag{A.5}$$

where $\bar{\boldsymbol{\rho}}_j = \boldsymbol{\rho}_j$ for $j = 1, \ldots, M$, $\bar{\boldsymbol{\rho}}_j = \widetilde{\boldsymbol{\rho}}_{j-M}$ for $j = M+1, \ldots, 2M$, $\bar{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_j$ for $j = 1, \ldots, M$, and $\bar{\boldsymbol{\phi}}_j = \boldsymbol{\phi}_{j-M}$ for $j = M+1, \ldots, 2M$. For the sake of notation simplicity, let $\varphi_{i,j} \equiv \varphi\left(\ln x_{i-1}; \bar{\boldsymbol{\rho}}_j\right)$ for $j = 1, \ldots, 2M$. Lemma 2.7 in Hwang and Ding (1997) implies that, if φ_{j_1} and φ_{j_2} are not sign-equivalent for $j_1 \in \{1, \ldots, 2M\}$ and $j_2 \in \{1, \ldots, 2M\}$, (A.5) holds if and only if $\boldsymbol{\phi}_0$, $\widetilde{\boldsymbol{\phi}}_0$, and $\bar{\boldsymbol{\phi}}_j$ jointly vanish for every $j \in \{1, \ldots, 2M\}$. Conditions C2 and C3 in Assumption 4 however preclude that possibility because they guarantee that there are no irrelevant limiting regimes. Although this means that φ_{j_1} and φ_{j_2} must be sign-equivalent, they must also come from different models; otherwise it would contradict C2 in Assumption 4. There thus exist $j_1 \in \{0,\dots,M\}$ and $j_2 \in \{M+1,\dots,2M\}$ such that φ_{j_1} and φ_{j_2} are sign-equivalent. Assumption 4 implies that (A.4) holds only if $\boldsymbol{\phi}_m = \widetilde{\boldsymbol{\phi}}_m$ and $\boldsymbol{\theta}_m = \widetilde{\boldsymbol{\theta}}_m$, $m = 1,\dots,M$ given that C1 rules out the permutation of regimes. It now remains to show that $\boldsymbol{\theta}_0$ uniquely maximizes the log-likelihood function $\mathcal{L}(\boldsymbol{\theta})$. Letting $\psi_i(\boldsymbol{\theta}_0) = x_i/\epsilon_i(\boldsymbol{\theta}_0)$ denote the true conditional duration process, one may rewrite, as in Lumsdaine (1996), the maximization problem as

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left[\mathcal{L}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta}_0)\right] = \max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbb{E}\left[\ln\frac{\psi_i(\boldsymbol{\theta}_0)}{\psi_{u,i}} - \frac{\psi_i(\boldsymbol{\theta}_0)}{\psi_{u,i}} - 1\right].$$

In addition, for any $y > 0$, $m(y) = y - \ln(y) \geq 0$, so that $\mathbb{E}\left[\ln\frac{\psi_i(\boldsymbol{\theta}_0)}{\psi_{u,i}} - \frac{\psi_i(\boldsymbol{\theta}_0)}{\psi_{u,i}}\right] \leq 0$.

Given that $m(y)$ achieves its maximum at $y = 1$, $\mathbb{E}[m(y)] \leq \mathbb{E}[m(1)]$ with equality holding almost surely only if $\ln\psi_i(\boldsymbol{\theta}_0)$ and $\ln\psi_{u,i}$ coincide with probability one. By the mean value theorem, this is equivalent to showing that

$$(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\,\frac{\partial\ln\psi_{u,i}}{\partial\boldsymbol{\theta}} = 0$$

with probability one. A straightforward application of Lemma 1 then shows that this happens if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, completing the proof. *Q.E.D.*

Proof of Theorem 4. We must verify whether the regularity conditions in Newey and McFadden (1994) hold under Assumptions 1 to 4. Assumption 3 takes care of their first condition, which relates to the compactness of the parameter space. Theorems 2 and 3 ensure the validity of their second and third conditions, which require the log-likelihood function to be continuous in the parameter vector $\boldsymbol{\theta}$, with a unique maximum at $\boldsymbol{\theta}_0$, and measurable with respect to $\{x_i,\ i \in \mathbb{N}\}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Finally, Lemma 2 fulfills the requirements for $\mathcal{L}_{u,N}(\boldsymbol{\theta}) \xrightarrow{p} \mathcal{L}(\boldsymbol{\theta})$. This means that $\widehat{\boldsymbol{\theta}}_{u,N} \xrightarrow{p} \boldsymbol{\theta}_0$, so that it now remains to demonstrate that $\left\|\widehat{\boldsymbol{\theta}}_N - \widehat{\boldsymbol{\theta}}_{u,N}\right\| \xrightarrow{p} 0$. We do that in Lemma 3 by showing that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_{u,N}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})| \xrightarrow{p} 0$, and hence $\widehat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}_0$. *Q.E.D.*

Proof of Theorem 5. As before, we first tackle the asymptotic normality of the QML estimator that hinges on $\mathcal{L}_{u,N}(\boldsymbol{\theta})$ and then employ Lemmas 3 and 5 to extend the result for the QML estimator based on $\mathcal{L}_N(\boldsymbol{\theta})$. Asymptotic normality of the QML estimator requires four additional regularity conditions. First, the true parameter vector $\boldsymbol{\theta}_0$ must lie in the interior of the parameter space $\boldsymbol{\Theta}$. Second, the matrix $\mathbf{A}_N(\boldsymbol{\theta})$ exists and is continuous in $\boldsymbol{\Theta}$. Third, the matrix $\mathbf{A}_N(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{A}_0$ for any sequence $\boldsymbol{\theta}_N$ such that $\boldsymbol{\theta}_N \xrightarrow{p} \boldsymbol{\theta}_0$. Fourth, the score vector satisfies

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{B}_0).$$

We next verify these conditions. Assumption 3 ensures that the first condition holds, whereas Lemma 5 substantiates the third condition. The second condition follows from the stationarity of the FC-LACD model and from the fact that $\ell_i(\boldsymbol{\theta})$ is twice differentiable on $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. In fact, $\mathbf{A}_0$ and $\mathbf{B}_0$ are nonsingular due to the model identifiability. Finally, Lemma 4 shows that the score condition also holds, completing the proof. *Q.E.D.*

Proof of Theorem 6. The local approximation to the instantaneous quasi-log-likelihood function in a neighborhood of $\mathcal{H}_0$ is $\ell_i(\boldsymbol{\theta}) = -\ln\psi_i(\boldsymbol{\theta}) - x_i/\psi_i(\boldsymbol{\theta})$. Let $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right]'$ with

$$\boldsymbol{\theta}_1 = \left(\tilde{\omega}_0, \omega_1, \dots, \omega_{M-1}, \tilde{\alpha}_0, \alpha_1, \dots, \alpha_{M-1}, \tilde{\beta}_0, \beta_1, \dots, \beta_{M-1}, c_1, \dots, c_{M-1}, \gamma_1, \dots, \gamma_{M-1}\right)'$$

and $\boldsymbol{\theta}_2 = (\delta_1, \delta_2)'$. The resulting score vector thus is

$$\mathbf{q}(\boldsymbol{\theta}) = (\mathbf{q}(\boldsymbol{\theta}_1)', \mathbf{q}(\boldsymbol{\theta}_2)')' = \sum_{i=1}^{N} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}_1} \ell_i(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \boldsymbol{\theta}_2} \ell_i(\boldsymbol{\theta}) \end{pmatrix} = \sum_{i=1}^{N} \left(\frac{x_i}{\psi_i} - 1 \right) \begin{pmatrix} \mathbf{v}_i \\ \mathbf{u}_i \end{pmatrix}$$

with $\mathbf{v}_i = \partial \ln \psi_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_1$ and $\mathbf{u}_i = \partial \ln \psi_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_2$. whereas the information matrix reads

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{1}{\psi_i^2} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}'} \frac{x_i}{\psi_i} - \left(\frac{x_i}{\psi_i} - 1 \right) \frac{\partial}{\partial \boldsymbol{\theta}'} \left(\frac{1}{\psi_i} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \right) \right] = \mathbb{E}\left[\frac{1}{\psi_i^2} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}} \frac{\partial \psi_i}{\partial \boldsymbol{\theta}'} \right] = \mathbb{E} \begin{bmatrix} \mathbf{v}_i \mathbf{v}_i' & \mathbf{v}_i \mathbf{u}_i' \\ \mathbf{u}_i \mathbf{v}_i' & \mathbf{u}_i \mathbf{u}_i' \end{bmatrix}.$$

Consider next the consistent estimator for the information matrix $\mathbf{A}(\boldsymbol{\theta})$ given by

$$\mathbf{A}_N(\boldsymbol{\theta}) = \frac{1}{2N} \sum_{i=1}^{N} \begin{pmatrix} \mathbf{v}_i \mathbf{v}_i' & \mathbf{v}_i \mathbf{u}_i' \\ \mathbf{u}_i \mathbf{v}_i' & \mathbf{u}_i \mathbf{u}_i' \end{pmatrix}$$

and let $\mathbf{d}_i = (\mathbf{v}_i', \mathbf{u}_i')'$. The LM statistic thus is

$$LM = \mathbf{q}(\boldsymbol{\theta})|_{\mathbb{H}_0} \left[\mathbf{A}_N(\boldsymbol{\theta})|_{\mathbb{H}_0} \right]^{-1} \mathbf{q}(\boldsymbol{\theta})|_{\mathbb{H}_0} = N \left[\sum_{i=1}^{N} \left(\frac{x_i}{\psi_i} - 1 \right) \mathbf{d}_i \right] \left(\sum_{i=1}^{N} \mathbf{d}_i \mathbf{d}_i' \right)^{-1} \left[\sum_{i=1}^{N} \left(\frac{x_i}{\psi_i} - 1 \right) \mathbf{d}_i \right].$$

To complete the proof, it then suffices to apply Lemmas 4 and 5. *Q.E.D.*

Proof of Theorem 7. The approximation error consists of

$$\begin{aligned} \ln \psi_i^{(N)} - \ln \psi_i &= \sum_{j=0}^{\infty} \left(\beta_N^j \bar{\omega}_{Nj} - \beta_0^j \bar{\omega}_{0j} \right) - \sum_{j=J_N+1}^{\infty} \beta_N^j \, \bar{\omega}_{Nj} \\ &= \sum_{j=0}^{\infty} \beta_0^j \left(\bar{\omega}_{Nj} - \bar{\omega}_{0j} \right) + \sum_{j=0}^{J_N} \left(\beta_N^j - \beta_0^j \right) \bar{\omega}_{Nj} - \sum_{j=J_N+1}^{\infty} \beta_0^j \, \bar{\omega}_{Nj}, \end{aligned}$$

where $\bar{\omega}_{0j} = \omega_0(\ln x_{i-1-j})$ and $\bar{\omega}_{Nj} = \omega_N(\ln x_{i-1-j})$. There are two approximation errors. The first is due to the finite number of regimes in the neural network (i.e., first two terms), whereas the second stems from the lag truncation (i.e., third term). Lemma 6 shows that the latter is at most of order $O_p\left(\beta^{J_N}\right)$. As for the former, Proposition 1 in Chen and Shen (1998) ensures that it is at most of order $O_p\left([N/\ln N]^{-1/3}\right)$ provided that $M_N^3 \ln M_N = O(N)$. We must then show that Conditions A.1 to A.4 in Chen and Shen (1998) hold within the semiparametric FC-LACD context. Definition 2 ensures geometric ergodicity given that $|\beta_0| < 1$, thereby satisfying Condition A.1, whereas Condition A.3 ensues exactly as in their proof of the Case 1.1 of Proposition 1. It remains to show that Conditions A.2 and A.4 are also valid. The former requires that $\sup_{\|\theta_N - \theta_0\| \leq \delta_1} \mathbb{V}(\ell_{\theta_N i} - \ell_{\theta_0 i}) \leq O(\delta_1^2)$ for all small $\delta_1 > 0$, whereas the latter necessitates that, for any $\delta_2 > 0$, there exists $0 < s < 2$ and a measurable function $U_N(\cdot)$ such that $\sup_{\|\theta_N - \theta_0\| \leq \delta_1} |\ell_{\theta_N i} - \ell_{\theta_0 i}| \leq \delta_2^s \, U_N(\ln x_i)$ with $\sup_N \mathbb{E}\left[U_N^{\delta_3}(\ln x_i) \right] \leq O(1)$ for $\delta_3 > 2$. Letting $\theta_N \equiv \sum_{i=1}^{J_N} \beta_N^j \omega_N(\ln x_{i-j-1})$, $\theta_0 \equiv \sum_{i=1}^{\infty} \beta_0^j \omega_0(\ln x_{i-j-1})$, $e_i(\theta_N) = \ln \epsilon_i(\theta_N)$ and $e_i(\theta_0) = \ln \epsilon_i(\theta_0) = \ln \epsilon_i$ yields

$$\begin{aligned} \ell_{\theta_N i} - \ell_{\theta_0 i} &= -\frac{1}{2} \left(e_i^2(\theta_N) - e_i^2(\theta_0) \right) = -\frac{1}{2} \left[e_i(\theta_N) - e_i(\theta_0) \right] \left[e_i(\theta_N) + e_i(\theta_0) \right] \\ &= (\theta_N - \theta_0) \left(e_i(\theta_0) - \frac{\theta_N - \theta_0}{2} \right). \end{aligned}$$

Condition A.4 then holds for $s = 2/3$ and $U_N(\ln x_i) = |e_i(\theta_0)| + c_N$, with $c_N \to \infty$ arbitrarily slowly, given that $\|\theta_N\| \leq c$ for every θ_N in the sieves parameter space and

$$|\ell_{\theta_N i} - \ell_{\theta_0 i}| \leq \|\theta_N - \theta_0\|_{\sup} \left[|e_i(\theta_0)| + \frac{\|\theta_N\|_{\sup} + \|\theta_0\|_{\sup}}{2} \right].$$

(Note the implicit use of Lemma 2 in Chen and Shen (1998) to change from the sup norm to the L^2 norm.) It now remains to bound the variance of $\ell_{\theta_N i} - \ell_{\theta_0 i}$. The first moment is

$$\mathbb{E}\left[\ell_{\theta_N i} - \ell_{\theta_0 i}\right] = \mathbb{E}\left[(\theta_N - \theta_0) \left(e_i(\theta_0) - \frac{\theta_N - \theta_0}{2} \right) \right] = -\frac{1}{2}\, \mathbb{E}(\theta_N - \theta_0)^2,$$

whereas the second moment reads

$$\begin{aligned} \mathbb{E}\left[\ell_{\theta_N i} - \ell_{\theta_0 i}\right]^2 &= \mathbb{E}\left(\theta_N - \theta_0\right)^2 \mathbb{E}\left[e_i^2(\theta_0)\right] + \frac{1}{4}\, \mathbb{E}\left(\theta_N - \theta_0\right)^4 = \sigma_\epsilon^2\, \mathbb{E}\left(\theta_N - \theta_0\right)^2 + \frac{1}{4}\, \mathbb{E}\left(\theta_N - \theta_0\right)^4 \\ &\leq \left\{ \sigma_\epsilon^2 + \mathbb{E}\left[\sup_{\|\theta_N - \theta_0\| \leq \delta_1} \left(\theta_N - \theta_0\right)^2 \right] \right\} \mathbb{E}\left(\theta_N - \theta_0\right)^2. \end{aligned}$$

Therefore, the variance is at most of order $O(\delta_1^2)$ as in Condition A.2 of Chen and Shen (1998). *Q.E.D.*

APPENDIX B. LEMMAS

LEMMA 1. *Let x_i follows a FC-LACD process with $M+1$ regimes given by (1)–(5) that satisfies Assumptions 1 to 4. Let $\mathbf{d}$ be a constant vector with the same dimension as $\boldsymbol{\theta}$. It then follows that*

$$\mathbf{d}' \left(\frac{\partial \ln \psi_{u,i}}{\partial \boldsymbol{\theta}} \right) = 0 \qquad a.s.$$

if and only if $\mathbf{d} = \mathbf{0}$.

PROOF. We follow the same reasoning as in the proof of Lemma 5 in Lumsdaine (1996). Define $\boldsymbol{\xi}_i \equiv \partial \ln \psi_i / \partial \boldsymbol{\theta}$ and $G_{m,i} \equiv G(\ln x_i; \gamma_m, c_m)$. It is straightforward to show that

$$\boldsymbol{\xi}_i = \beta(\ln x_{i-1}) \boldsymbol{\xi}_{i-1} + \boldsymbol{\kappa}_{i-1},$$

where

$$\begin{aligned} \boldsymbol{\kappa}_{i-1} = \Big[&1,\, \ln x_{i-1},\, \ln \psi_{i-1},\, G_{1,i-1},\, \ldots,\, G_{M,i-1}, \\ &G_{1,i-1} \ln x_{i-1},\, \ldots,\, G_{M,i-1} \ln x_{i-1},\, G_{1,i-1} \ln \psi_{i-1},\, \ldots,\, G_{M,i-1} \ln \psi_{i-1}, \\ &(\omega_1 + \alpha_1 \ln x_{i-1} + \beta_1 \ln \psi_{i-1}) \frac{\partial G_{1,i-1}}{\partial \gamma_1},\, \ldots, \\ &(\omega_M + \alpha_M \ln x_{i-1} + \beta_M \ln \psi_{i-1}) \frac{\partial G_{M,i-1}}{\partial \gamma_M}, \\ &(\omega_1 + \alpha_1 \ln x_{i-1} + \beta_1 \ln \psi_{i-1}) \frac{\partial G_{1,i-1}}{\partial c_1},\, \ldots, \\ &(\omega_M + \alpha_M \ln x_{i-1} + \beta_M \ln \psi_{i-1}) \frac{\partial G_{M,i-1}}{\partial c_M} \Big]', \end{aligned}$$

so that $\mathbf{d}'\boldsymbol{\xi}_i = \mathbf{d}'\beta(\ln x_{i-1})\boldsymbol{\xi}_{i-1} + \mathbf{d}'\boldsymbol{\kappa}_{i-1}$. It then follows by assumption that $\mathbf{d}'\boldsymbol{\xi}_i = 0$ and $\mathbf{d}'\boldsymbol{\xi}_{i-1} = 0$ with probability one, implying that $\mathbf{d}'\boldsymbol{\kappa}_{i-1} = 0$ with probability one. In view that $\boldsymbol{\kappa}_i$ is nondegenerate, $\mathbf{d}'\boldsymbol{\xi}_i = 0$ with probability one if and only if $\mathbf{d} = \mathbf{0}$. *Q.E.D.*

LEMMA 2. *If x_i follows a FC-LACD process with $M+1$ regimes given by (1)–(5) that satisfies Assumptions 1 to 4, then* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_{u,N}(\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})| \xrightarrow{p} 0$.

PROOF. We derive this result by building on the proof of Lemma 4.3 in Ling and McAleer (2003). Let $g(\mathbf{X}_i, \boldsymbol{\theta}) = \ell_{u,i}(\boldsymbol{\theta}) - \mathbb{E}\left[\ell_{u,i}(\boldsymbol{\theta})\right]$, where $\mathbf{X}_i = (x_i, x_{i-1}, x_{i-2}, \ldots)'$. Theorem 2 implies that $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |g(\mathbf{X}_t, \boldsymbol{\theta})|\right] < \infty$. The result then ensues from the fact that Theorem 3.1 in Ling and McAleer (2003) implies that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left|N^{-1}\sum_{i=1}^N g(\mathbf{X}_i, \boldsymbol{\theta})\right| = o_p(1)$ in view that $g(\mathbf{X}_t, \boldsymbol{\theta})$ is stationary with zero mean. *Q.E.D.*

LEMMA 3. *If x_i follows a FC-LACD process with $M+1$ regimes given by (1)–(5) that satisfies Assumptions 1 to 4, then* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_{u,N}(\boldsymbol{\theta}) - \mathcal{L}_N(\boldsymbol{\theta})| \xrightarrow{p} 0$.

PROOF. We follow the proof of the first result in Lemma 6 in Lumsdaine (1996). The conditions in Theorem 1 ensure that $\ln\psi_{u,0}$ is well defined and that, as the constant $k \to \infty$,

$$\Pr\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} (\ln\psi_{u,0}) > k\right] \to 0.$$

Combining (7) and (8) gives way to

$$\ln\psi_{u,i} - \ln\psi_i = (\ln\psi_{u,0} - \ln\psi_0)\prod_{j=1}^{i}\beta(\ln x_j).$$

Defining two finite positive constants $\underline{\delta}$ and $\bar{\delta}$ such that $\ln\psi_i > \underline{\delta}$ and $\beta(\ln x_i) \le \bar{\delta}$ then leads to

$$0 \le \left|N^{-1/2}\sum_{i=1}^{N}\ln\left(\frac{\psi_{u,i}}{\psi_i}\right)\right|^p \le \left[N^{-1/2}\sum_{i=1}^{N}\left|\ln\left(\frac{\psi_{u,i}}{\psi_i}\right)\right|\right]^p = \left[N^{-1/2}\sum_{i=1}^{N}\ln\left(\frac{\psi_{u,0}}{\psi_0}\right)\prod_{j=1}^{i}\beta(\ln x_j)\right]^p$$
$$\le N^{-p/2}\left[\ln\left(\frac{\psi_{u,0}}{\underline{\delta}}\right)\right]^p\left[\sum_{i=1}^{N}\bar{\delta}^i\right]^p.$$

The upper bound of the latter expression converges in probability uniformly to zero by Theorem 1 and Slutsky's Theorem, and hence $\Pr\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{i=1}^N |\ln\psi_{u,i} - \ln\psi_i| > k\right] \to 0$ as the sample size grows for any constant $k > 0$. It remains to show that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|N^{-1/2}\sum_{i=1}^N\left(\frac{x_i}{\psi_{u,i}} - \frac{x_i}{\psi_i}\right)\right| \xrightarrow{p} 0$. To that end, we first observe that

$$\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left|x_i\left(\frac{\psi_i - \psi_{u,i}}{\psi_{u,i}\psi_i}\right)\right|\right]^p \le \frac{1}{N^{p/2}\underline{\delta}^{2p}}\left[\sum_{i=1}^{N}|x_i(\psi_i - \psi_{u,i})|\right]^p = \frac{1}{N^{p/2}\underline{\delta}^{2p}}\left[\sum_{i=1}^{N}x_i^2\left|(\psi_0 - \psi_{u,0})\prod_{j=1}^{i}\beta(\ln x_j)\right|\right]^p.$$

Define $\xi_i \equiv \left|(\psi_0 - \psi_{u,0})\prod_{j=1}^{i}\beta(\ln x_j)\right|$. Under the conditions of Theorem 1, the duration process is a strictly stationary and ergodic with $\mathbb{E}|x_i| < \infty$. In addition, it holds that $\sup_{1\le i\le N}|\xi_i| \le C_N$, where C_N is some finite constant, and $\frac{1}{N}\sum_{i=1}^N|\xi_i| = o_p(1)$. To conclude the proof, it now suffices to apply Lemma 4.5 in Ling and McAleer

(2003) to show that $N^{-1}\sum_{i=1}^{N} x_i\xi_i = o_p(1)$, and hence $\frac{1}{N^{p/2}\underline{\delta}^{2p}}\left[\sum_{i=1}^{N} x_i \left|(\psi_0 - \psi_{u,0})\prod_{j=1}^{i}\beta(\ln x_j)\right|\right]^p \xrightarrow{p} 0$. *Q.E.D.*

LEMMA 4. *The conditions of Theorem 5 ensure not only that* $\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\,\ell_i(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]$ *exists and is finite, but also that* $\mathbf{B}_0$ *is finite and positive definite, and that*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{B}_0).$$

PROOF. The existence of $\mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\,\ell_i(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] < \infty$ immediately follows from the conditions of Theorem 1. Letting then

$$\nabla_0\,\ell_{u,i} \equiv \frac{\partial\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \qquad \text{and} \qquad \nabla_0 \ln\psi_{u,i} \equiv \frac{\partial\ln\psi_{u,i}}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

yields $\nabla_0\,\ell_{u,i}\,\nabla_0\,\ell'_{u,i} = \left(\epsilon_i^2 - 2\epsilon_i + 1\right)\nabla_0\ln\psi_{u,i}\,\nabla_0\ln\psi'_{u,i}$. Let $\underline{\delta} < \infty$ be a positive constant such that $\ln\psi_{u,i} > \underline{\delta}$. Under the strict stationarity condition of Theorem 1, we may employ the same reasoning as in the proof of Lemma 1 in Boussama (2000), giving way to $\mathbb{E}\left[\nabla_0\ln\psi_{u,i}\,\nabla_0\ln\psi'_{u,i}\right] < \mathbf{K}_1$, where $\mathbf{K}_1$ is a constant matrix with finite elements, and $\mathbb{E}\left[\nabla_0\,\ell_{u,i}\,\nabla_0\,\ell'_{u,i}\right] \le \mathbf{K}_1\,\mathbb{E}\left(\epsilon_i^2 - 2\epsilon_i + 1\right) = \mathbf{K}_1\,(\mu_2 - 1)$, which is finite given that $\mu_2 \equiv \mathbb{E}\left(\epsilon_i^2\right) < \infty$. This means that $\mathbf{B}_0$ is finite. The conditions of Theorems 1 and 3 also ensure that $\mathbf{B}_0$ is positive definite. It now remains to show the asymptotic normality of the score vector. Let $S_N = \sum_{i=1}^{N}\mathbf{k}'\nabla_0\,\ell_{u,i}$, where $\mathbf{k}$ is a constant vector. It then follows that S_N is a martingale with respect to the filtration $\mathcal{F}_i$ with a positive expected value, and hence $N^{-1/2}S_N \xrightarrow{d} \mathcal{N}\left(0,\mathbf{k}'\mathbf{B}_0\mathbf{k}\right)$ by the central limit theorem in Stout (1974). A straightforward application of the Cramér-Wold device then yields $N^{-1/2}\sum_{i=1}^{N}\frac{\partial\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} \mathcal{N}\left(0,\mathbf{B}_0\right)$, whereas it is also possible to show, as in the proof of Lemma 3, that

$$N^{-1/2}\sum_{i=1}^{N}\left|\frac{\partial\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} - \frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right| \xrightarrow{p} \mathbf{0} \text{ and } N^{-1/2}\sum_{i=1}^{N}\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{d} \mathcal{N}(0,\boldsymbol{B}_0),$$

completing the proof. *Q.E.D.*

LEMMA 5. *Under the conditions of Theorem 5, both*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \mathbb{E}\left[\frac{\partial^2\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right| \quad \textit{and} \quad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial^2\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right|$$

converge in probability to zero.

PROOF. If we define

$$\nabla_0^2\,\ell_{u,i} \equiv \frac{\partial^2\ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \qquad \text{and} \qquad \nabla_0^2\ln\psi_{u,i} \equiv \frac{\partial^2\ln\psi_{u,i}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

then

$$\nabla_0^2\,\ell_{u,i} = \left(\frac{x_i}{\psi_{u,i}} - 1\right)\nabla_0^2\ln\psi_{u,i} - \left(\frac{x_i}{\psi_{u,i}}\right)\nabla_0\ln\psi_{u,i}\nabla_0\ln\psi'_{u,i}.$$

Because $\nabla_0^2\ln\psi_{u,i}$ consists exclusively of second-order terms, it ensues from the conditions of Theorem 1 that $\mathbb{E}\left[\nabla_0^2\ln\psi_{u,i}\right] \le \mathbf{K}_2$, where $\mathbf{K}_2$ is a constant matrix with finite elements. This implies that

$$\nabla_0^2\ell_{u,i} \le (\epsilon_i - 1)\,\mathbf{K}_2 - \epsilon_i\,\mathbf{K}_1,$$

so that $\mathbb{E}\left[\nabla_0^2 \ell_{u,i}\right]$ is finite. By Theorem 3.1 in Ling and McAleer (2003), $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 \ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \mathbb{E}\left[\frac{\partial^2 \ell_{u,i}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right| \xrightarrow{p} 0$. We omit the proof of the second result given that it very much resembles the proof of Lemma 3. *Q.E.D.*

LEMMA 6. *If the semiparametric FC-LACD process is stationary with finite second moments and $\beta-$mixing with exponential decay, the approximation error due to the lag truncation is negligible in that*

$$\sum_{j=J_N+1}^{\infty} \beta_0^j \,\bar\omega_{Nj} \le O_p\left(\beta_0^{J_N}\right). \tag{B.6}$$

PROOF. We first show that both the expectation and variance of the left-hand side of (B.6) converge to zero as the sample size increases and then complete the proof applying Chebyshev's inequality. Stationarity implies that

$$\mathbb{E}\left(\sum_{j=J_N+1}^{\infty} \beta_0^j \,\bar\omega_{Nj}\right) = \mathbb{E}\left(\bar\omega_{N1}\right) \sum_{j=J_N+1}^{\infty} \beta_0^j = \mathbb{E}\left(\bar\omega_{N1}\right) \frac{\beta_0^{J_N+1}}{1-\beta_0}$$

and that $\mathbb{V}\left(\sum_{j=J_N+1}^{\infty} \beta_0^j \,\bar\omega_{Nj}\right) = \sum_{j=J_N+1}^{\infty} \beta_0^{2j}\,\mathbb{V}\left(\bar\omega_{Nj}\right) + 2 \sum\sum_{J_N+1\le j<k<\infty} \beta_0^{j+k} \text{Cov}\left(\bar\omega_{Nj}, \bar\omega_{Nk}\right)$. The first term of the right-hand side is of order $O\left(\beta_0^{2J_N}\right)$. As for the second term, we take benefit from the fact that the semiparametric FC-LACD process is, by assumption, $\beta-$mixing with exponential decay to employ the covariance inequality in Lemma A1 in Dette and Spreckelsen (2004): for some $\delta > 0$, $|\text{Cov}(\bar\omega_{Nj}, \bar\omega_{Nk})| \le 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}} \rho_{k-j}^{\frac{\delta}{1+\delta}}$, $1 \le j < k$, where ρ_j is the mixing coefficient; see also Lemma 1 in Yoshihara (1976). In particular, exponential decay means that the latter is of order $O(\eta^j)$ with $0 < \eta < 1$ and so

$$\begin{aligned}
\left|\sum\sum_{J_N+1\le j<k<\infty} \beta_0^{j+k}\,\text{Cov}(\bar\omega_{Nj}, \bar\omega_{Nk})\right| &\le 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}} \sum\sum_{J_N+1\le j<k<\infty} \beta_0^{j+k}\, O\left(\eta^{\frac{\delta}{1+\delta}(k-j)}\right) \\
&= 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}} \sum_{j=J_N+1}^{\infty} \beta_0^j\, O\left(\eta^{-\frac{\delta}{1+\delta}\,j}\right) \sum_{k=j+1}^{\infty} \beta_0^k\, O\left(\eta^{\frac{\delta}{1+\delta}\,k}\right) \\
&= 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}} \sum_{j=J_N+1}^{\infty} \beta_0^j\, O\left(\eta^{-\frac{\delta}{1+\delta}\,j}\right) \frac{\beta_0^{j+1}\, O\left(\eta^{\frac{\delta}{1+\delta}\,(j+1)}\right)}{1-\beta\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)} \\
&= 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}} \sum_{j=J_N+1}^{\infty} \beta_0^{2j+1}\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\left[1-\beta\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\right]^{-1} \\
&= 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}}\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\left[1-\beta\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\right]^{-1} \sum_{j=J_N+1}^{\infty} \beta_0^{2j+1} \\
&= 4\left[\mathbb{E}|\bar\omega_{N1}^{1+\delta}|\right]^{\frac{1}{1+\delta}}\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\left[1-\beta\, O\left(\eta^{\frac{\delta}{1+\delta}}\right)\right]^{-1} \frac{\beta_0^{2J_N+3}}{1-\beta_0^2} = O\left(\beta_0^{2J_N}\right).
\end{aligned}$$

This means that the expectation and standard deviation of (B.6) are both at most of order $O\left(\beta_0^{J_N}\right)$, and hence the approximation error due to the lag truncation is at most of order $O_p\left(\beta_0^{J_N}\right)$ by Chebyshev's inequality. *Q.E.D.*

References

Bauwens, L., Giot, P., 2000, The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks, Annales d'Economie et de Statistique 60, 117–150.

Boussama, F., 2000, Normalité asymptotique de l'estimateur du pseudo-maximum de vraisemblance d'un modèle GARCH, Comptes Rendus de l'Academie des Sciences, Série I 331, 81–84.

Chen, R., Tsay, R., 1991, On the ergodicity of TAR(1) processes, Annals of Applied Probability 1, 613–634.

Chen, X., Shen, X., 1998, Sieve extremum estimates for weakly dependent data, Econometrica 66, 289–314.

Dette, H., Spreckelsen, I., 2004, Some comments on specification tests in nonparametric absolutely regular processes, Journal of Time Series Analysis 25, 159–172.

Drost, F., Werker, B. J. M., 2004, Semiparametric duration models, Journal of Business and Economic Statistics 22, 40–50.

Engle, R. F., Russell, J. R., 1998, Autoregressive conditional duration: A new model for irregularly-spaced transaction data, Econometrica 66, 1127–1162.

Fernandes, M., Grammig, J., 2006, A flexible family of autoregressive conditional duration models, Journal of Econometrics 130, 1–23.

Grenander, U., 1981, Abstract Inference, Wiley Series, New York.

Hornik, K., Stinchcombe, M. B., White, H., 1989, Multilayer feedforward networks are universal approximators, Neural Networks 2, 359–366.

Hwang, J. T. G., Ding, A. A., 1997, Prediction intervals for artificial neural networks, Journal of the American Statistical Association 92, 109–125.

Ling, S., 1999, On the probabilistic properties of a double-threshold arma conditional heteroskedastic model, Journal of Applied Probability 36, 688–705.

Ling, S., McAleer, M., 2003, Asymptotic theory for a vector ARMA-GARCH model, Econometric Theory 19, 280–310.

Lumsdaine, R. L., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica 64, 575–596.

Makovoz, Y., 1996, Random approximants and neural networks, Journal of Approximation Theory 85, 98–109.

Medeiros, M. C., Veiga, A., 2000, A hybrid linear-neural model for time series forecasting, IEEE Transactions on Neural Networks 11, 1402–1412.

Medeiros, M. C., Veiga, A., 2009, Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model, Econometric Theory 25, 117–161.

Meitz, M., 2006, A necessary and sufficient condition for the strict stationarity of a family of GARCH processes, Econometric Theory 22, 985–988.

Meitz, M., Saikkonen, P., 2008, Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models, Econometric Theory 24, 1291–1320.

Meitz, M., Teräsvirta, T., 2006, Evaluating models of autoregressive conditional duration, Journal of Business and Economic Statistics 24, 104–124.

Newey, W., McFadden, D. L., 1994, Large sample estimation and hypothesis testing, in: R. F. Engle D. L. McFadden (eds), Handbook of Econometrics, Vol. 4, Elsevier Science, pp. 2111–2245.

Table 1
Descriptive statistics of IBM price durations

We document the descriptive statistics for both plain and diurnally adjusted durations for the period running from September to November 1996. The in-sample period considers the first 4,484 observations of the data set, whereas the remaining 2,244 observations compose the out-of-sample period.

	adjusted durations		plain durations	
	in-sample	out-of-sample	in-sample	out-of-sample
sample size	4,484	2,244	4,484	2,244
mean	1.2387	0.5682	262.55	119.74
standard deviation	1.6470	0.7541	422.67	172.96
minimum	0.0039	0.0033	1	1
first quartile	0.2902	0.1540	51	30
median	0.7137	0.3236	128	60
third quartile	1.5399	0.6744	300	139
maximum	29.121	11.286	7,170	2,865
overdispersion	1.3296	1.3271	1.6098	1.4445

Stout, W. F., 1974, Almost Sure Convergence, Academic Press, New York.

Suarez-Fariñas, M., Pedreira, C. E., Medeiros, M. C., 2004, Local-global neural network: A new approach for nonlinear time series modeling, Journal of the American Statistical Association 99, 1092–1107.

Tjøstheim, D., 1990, Non-linear time series and markov chains, Advances in Applied Probability 22, 587–611.

Tweedie, R. L., 1975, Sufficient conditions for ergodicity and recurrence of markov chains on a general state space, Stochastic Processes and Their Applications 3, 385–403.

White, H., 1996, Estimation, Inference and Specification Analysis, Cambridge University Press, New York.

Wooldridge, J. M., 1990, A unified approach to robust, regression-based specification tests, Econometric Theory 6, 17–43.

Yoshihara, K.-i., 1976, Limiting behavior of U-statistic for stationary, absolutely regular processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 35, 237–252.

Zhang, M. Y., Russell, J. R., Tsay, R. S., 2001, A nonlinear autoregressive conditional duration model with applications to financial transaction data, Journal of Econometrics 104, 179–207.

Table 2

Estimation results for IBM price durations

We document estimation results for diurnally adjusted durations from September to November 199. The model is estimated by quasi-maximum likelihood for the first 4,484 observations of the sample. The lower panel reports the p-values of the LM test for an extra regime, and the in-sample and out-of-sample values of the logarithm of the quasi-likelihood function.

		one regime		two regimes	
		estimate	standard error	estimate	standard error
ω_0		0.0501	0.0015	0.0201	0.0028
ω_1				0.0152	0.0362
α_0		0.0867	0.0040	0.0609	0.0132
α_1				0.1118	0.0158
β_0		0.8929	0.0043	0.9301	0.0062
β_1				-0.2331	0.0290
c_1				0.3210	0.0160
γ_1				496.99	0.0004
LM test for extra regime (p-value)			0.0000		0.3765
log-likelihood	in-sample		-1.1247		-1.1172
	out-of-sample		-0.3788		-0.3732